Tests non paramétriques sous Microsoft® Excel

è télécharger les tests - récupérer le fichier au format.pdf -



revu le 3 juillet 2003 - section en cours

Plan : I. **Principaux tests non-paramétriques** : 1. chi2 - 2. tableau de contingence - test exact de Fisher - G-Test ou Log-Likelihood ratio - 3. Mann-Whitney - 4. Fisher-Yates-Terry - 5. T de Wilcoxon - Sign Test - 6. kruskal-Wallis - 7. Wald-Wolfowitz - run test - Moses Test of Extreme Reaction - 8. Coefficient de Spearman - 9. test de Friedman 9bis. test de Page - 9ter - coefficient de concordance W de Kendall - 10. test de Kolmogorov Smirnov - 11. Tau de Kendall - 12. test de la médiane [Mood test - Extension of the median Test] 13. test de McNemar - 14. test de Jonckheere Terpstra - II. **autres tests** : Durbin-Watson - III. le logiciel : feuille - commandes - boîte de dialogue du menu - Bibliographie [références par tests, cf. infra]- Tables - Liens - logiciels -

Remerciements: je dois à M. Christian Delaruelle d'avoir mis en évidence un bogue majeur dans le calcul des rangs sur le test de Mann Whitney. Non seulement ce bogue a été corrigé, mais cela m'a permis de repenser entièrement le mode de détermination du test de kruskal-Wallis dont la méthode de Mann Whitney ne constitue qu'un cas particulier pour deux colonnes. La méthode de calcul des rangs est maintenant basée sur l'élimination progressive des minima (ou du minimum quand il n'y a pas d'ex aequo) des colonnes de données. De substantielles améliorations ont été apportées dans la présentation des résultats des tests de Mann Whitney Wilcoxon [les deux sont liés intrinsèquement] et de kruskal Wallis. Consultez aussi le superbe site « ***Disciplus simplex*** » où l'on peut trouver de très nombreuses pages sur Excel. Et des meilleures. Cet algorithme peut servir pour de nombreux exemples où le calcul des rangs est nécessaire. Je rappelle que ces pages sur les statistiques n'ont de vertu que didactique et qu'elles sont proposées - à titre d'illustration et de même que les programmes macro qui y sont annexées - pour faire valoir les vertus du raisonnement par analogie. Elles tâchent d'illustrer ce que l'on peut réaliser tout en ne comprenant rien aux mathématiques, pourvu que l'on ait un peu de bonne volonté et de l'obstination. En définitive, le Mail que m'a envoyé M. Delaruelle m'a conduit à repogrammer entièrement les tests non paramétriques et à faire de multiples addenda et errata.

Au fil de l'exposé, je commente le contenu de nombreux sites web sur les tests non paramétriques, en donnant bien sûr toutes les sources, car j'en tire des équations et des exemples. Je dois faire une mention spéciale pour un site unique où l'on trouve des tables ailleurs parfaitement introuvables :

http://www.dsa.unipr.it/soliani/soliani.html. Il s'agit de l'ensemble de cours en ligne le plus complet qui se puisse trouver sur le web, du moins à ma connaissance. Même si l'on ne parle pas l'italien, une oreille latine ne peut manquer de comprendre, par intuition immédiate, le texte généreusement proposé par l'auteur. De nombreux exemples viennent agrémenter les chapitres où l'étudiant trouvera toutes les formules mathématiques et toutes les tables voulues.

Nos macros sont - en principe - compatibles au travers des différentes versions - d'**Excel 4.0** à **Excel XP**.

Les tests non-paramétriques sont utilisés dès que l'effectif N d'un échantillon est inférieur à 30. On admet en effet, en biologie humaine tout au moins, que la normalité d'un échantillon est atteinte à partir de N =30. En dessous de N=30, les tests nécessitent certaines hypothèses [normalité des distributions, égalité des variances, etc.]. Ceci est particulièrement vrai quand les effectif sont très faibles. Une mention spéciale, à cet égard, est à décerner au très utile test de Kolmogorov Smirnov pour un échantillon.

« *Un test non paramétrique est donc un test d'hypothèse pour lequel il n'est pas nécessaire de spécifier la forme de la distribution de la population étudiée. Il faut cependant en général que les observations soient indépendantes, c'est-à-dire que la sélection d'un quelconque individu dans la population en vue de former l'échantillon ne doit pas influencer le choix des autres individus.*
Les méthodes non paramétriques requièrent peu d'hypothèses concernant la population étudiée. Elles ignorent notamment l'hypothèse classique de la normalité de la population.
Ces tests peuvent être appliquer, comme nous l'avons dit, à de petits échantillons.
Ils peuvent s'appliquer à des caractères qualitatifs, à des grandeurs de mesure, à des rangs de classement. Ils peuvent s'appliquer à des données incomplètes ou imprécises.
Des recherches théoriques ont montré que l'efficacité des tests non paramétriques n'est que légèrement inférieure à celle de leurs équivalents paramétriques quand la distribution de la population étudiée est spécifiée, par exemple la loi normale. Elle est en revanche supérieure à celle des tests paramétriques quand la distribution de la population dévie sensiblement de la distribution spécifiée (normale). » [Michel Le-Her.]

Il convient de s'assurer de la normalité d'un échantillon, avant d'employer un test dit paramétrique, comme le test t de Student Fisher.

Le test de Kolmogorov Smirnov, que nous retrouverons plus loin, peut encore être utilisé à cette fin [c'est le test le plus protéiforme que je connaisse ! Il permet de tester la normalité d'un échantillon, en le testant à travers différentes lois : uniforme, normale, Poisson, etc.]

« *Dans le cas général, le test de Kolmogorov-Smirnov vise à déterminer si les fonctions de répartition de deux populations sont identiques. Il est utilisé lorsqu'on est en présence de deux échantillons provenant de deux populations pouvant être différentes. Contrairement au test de Mann-Whitney ou au test de Wilcoxon [...] dont l'objet est de détecter des différences entre deux moyennes ou médianes, le test de Kolmogorov-Smirnov a l'avantage de prendre en considération les fonctions de répartitions dans leur ensemble.*
Mais le test de Kolmogorov-Smirnov peut aussi être utilisé comme test d'adéquation. Dans ce cas, on est en présence d'un seul échantillon aléatoire tiré d'une population; sa fonction de répartition sera comparée à une fonction de répartition spécifique et connue (par exemple celle de la loi normale, celle de la loi uniforme ou encore celle de la loi de chi-carré). » [Michel Le-Her.]

On pourra lire une documentation très précise, complète et pratique, sur l'emploi des tests non paramétriques et leur degré de pertinence, comparés à des tests paramétriques, sur le site : Cours de DEUG, Probabilités et Statistiques, Avner Ba-Hen, Aix-Marseille III.

I. Principaux tests non-paramétriques

1)- le chi2

La 1ère méthode est le test du chi2 ; l'inconvénient de ce test consiste en une perte d'informations par la transformation d'une variable quantitative en variable qualitative. Il est donc à déconseiller dans les cas où l'on peut utiliser une variable quantitative. Ce test est disponible en standard dans EXCEL. Voici un exemple :

	55,58
59	55,58
58	55,58
49	55,58
55	55,58
52	55,58
54	55,58
41	55,58
50	55,58
52	55,58
66	55,58
79	55,58
52	55,58

TABLEAU I
(chi2)

Sur cet exemple, on voit 2 colonnes, la 1ère représente la variable d'un échantillon de 12 individus ; la 2ème colonne est la moyenne [soit 55,48]. Le test du chi2 montre que la moyenne observée ne diffère pas de la moyenne théorique , avec p = 0,08 [non significatif : NS]. Toutefois, Excel ne permet que de comparer un vecteur d'observations avec un autre vecteur donnant les valeurs théoriques [ici, nous avons choisi la moyenne]. La formule à employer est :

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - EXP_i)^2}{EXP_i} \quad (1)$$

Dans l'exemple présenté, répétons-le, il s'agit de comparer une distribution observée à une distribution théorique : il s'agit du test du chi2 dit d'ajustement. Au §2, on envisage le chi2 dit d'homogénéité, permettant de mesurer le fait que deux échantillons appartiennent ou non à la même population.

2)- le tableau de contingence è ne figure pas dans l'*Analysis Tool Pack* d'Excel

« Another useful way of looking at the relationship between two nominal (or categorical) variables is to cross-classify the data and get a *count* of the number of cases sharing a given combination of levels (*i.e.*, categories), and then create a contingency table (*cross-tabulation*) showing the levels and the counts.

A *contingency table* lists the frequency of the joint occurrence of two levels (or possible outcomes), one level for each of the two categorical variables. The levels for one of the categorical variables correspond to the columns of the table, and the levels for the other categorical variable correspond to the rows of the table. The primary interest in constructing contingency tables is usually to determine whether there is any association (in terms of statistical dependence) between the two categorical variables, whose counts are displayed in the table. A measure of the global association between the two categorical variables is the *Chi-square* statistic, which is computed as follows:

Consider a contingency table with k rows and h columns. Let n_{ij} denote the cross-frequency of cell (i, j). Let H_{ij} denote the expected frequency of the cell. The deviation between the observed and expected frequencies characterizes the disagreement between the observation and the hypothesis of independence. The expected frequency for any cell can be calculated by the following formula:

$H_{ij}=(RT´\ CT)\ /\ N$

where

ñ. $_3$= expected frequency in a given cell (i, j)

RT = row total for the row containing that cell.

CT = column total for the column containing that cell.

N = total number of observations.

All the deviations can be studied by computing the quantity, denoted by

$$\chi^2 = \frac{\sum_i \sum_j (n_{ij} - \hat{n}_{ij})^2}{n_{ij}} \quad (2)$$

This statistic is distributed according to ***Pearson's Chi-square*** law with (k-1)× (h-1) degrees of freedom. Thus, the statistical significance of the relationship between two categorical variables is tested by using the test which essentially finds out whether the observed frequencies in a distribution differ significantly from the frequencies, which might be expected according to a certain hypothesis (say the hypothesis of independence between the two variables). » [*Guide to Advanced Data Analysis using IDAMS Software*, P.S. NAGPAUL, New Delhi (India)]

Ainsi, le test du chi2 élargi au tableau de contingence consiste à étudier plusieurs répartitions observées, quand on ne dispose pas de répartition théorique. Voici l'exemple de 3 répartitions observées dont on se demande si elles diffèrent :

A	88	24	27	61	20	25
B	63	17	20	39	27	25
C	41	15	18	22	31	17

	A	B	C	nij
1	0.59	0	0.93	1.52
2	0.01	0.11	0.09	0.21
3	0.01	0.09	0.21	0.31
4	1.74	0.03	2.27	4.04
5	5.09	0.07	6.99	12.15
6	0.39	0.39	0.01	0.79
nji	7.83	0.69	10.5	19.02
	Contingency table			
	ddl	10		
	chi2	19.01		
	p	0.04		

TABLEAU II
(tableau de contingence)

La réponse est positive puisque p = 0,04 pour 10 degrés de liberté [(colonnes-1) x (lignes-1)].

La dernière version du logiciel inclue la correction de Yates, la formulation uni ou bilatérale et le **test exact de Fisher** pour les tableaux à 4 cases avec effectifs théoriques < 5 [Cf. Fisher R. A. ***The logic of scientific inference***, Journal of the Royal Statistical Society, vol. 98, pp. 39-54]. Sur la correction de Yates, cf. ***Contingency tables involving small numbers and the*** c^2 ***test*** (Journal of the Royal Statistical Society Suppl. 1: 217-235). Voici d'abord un tableau 2 x 2 traité sans la correction de continuité de Yates :

15 9
10 16

Contingency table
ddl 1
chi2 2.885
p 0.089

et avec la correction :

$$\chi^2 = \frac{\sum_i \sum_j \left(\left| n_{ij} - n_{ji} \right| - \frac{1}{2} \right)^2}{n_{ji}}$$ (2bis)

Contingency table
ddl 1
chi2 2.003
p cor. Yates 0.157

Un autre auteur, Haber, a proposé une autre correction, mais qui n'est pas sensiblement différente de celle de Yates [***comparison of some continuity corrections for the chi-squared test on 2 x 2 tables***, Journal of the American Statistical

Association, vol. 75, pp. 510-515]. Voici à présent un exemple d'emploi d'un test exact de Fisher.

Soit le tableau à 4 cases :

	A	**B**
+	20	0
-	5	6

« *On considère une table de contingence 2 x 2. On note A et B les deux variables qualitatives observées sur n individus. Lorsque les effectifs sont trop petits, on transforme l'inconvénient des échantillons de petite taille en bénéfice en énumérant l'ensemble des arrangements possibles des observations puis en calculant les probabilités exactes de chaque arrangement.* » [D. Chessel, A.B. Dufour & J. Thioulouse - ***Biométrie et Biologie Evolutive*** - Université Lyon1]

Dans le cas présent, imaginons que l'on administre un médicament à deux groupes de malades A et B : dans le premier groupe, on obtient 20 succès [+] et 0 dans le groupe B. On ne peut pas appliquer le test du chi2 même en employant la correction de Yates, car l'un des effectifs théoriques est trop petit [pour la valeur **0** : **2.13**]. L'hypothèse H_0 s'énonce comme : A et B sont indépendantes. L'hypothèse alternative H_1 : A et B sont liées. Voici la marche à suivre pour les calculs :

a)- on détermine la quantité suivante :

$$\frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{n!\,a!\,b!\,c!\,d!} \quad (3)$$

où dans l'exemple choisi : **a = 20 ; b = 0 ; c = 5 et d = 6. n = 31**.

b)- on calcule ensuite les probabilités des tables de contingence présentant des situations aussi extrêmes que celle observée : p_j avec $j = 1,m$. Voyons cela sur notre exemple :

Possible matrices and their Ps

p for 0 : [20,**0**,5,6]
0.00062748
p for 1 : [19,**1**,6,5]
0.01254956
p for 2 : [18,**2**,7,4]
0.0851577
p for 3 : [17,**3**,8,3]
0.25547311
p for 4 : [16,**4**,9,2]
0.36192025

p for 5 : [15,**5**,10,1]
0.23162896
p for 6 : [14,**6**,11,0]
0.05264294

Il est intuitif de considérer qu'il existe 6 possibilités, avec des probabilités s'écartant plus ou moins de notre disposition, qui est celle correspondant à : **p for 0 = 0.00062748**. On remarque que l'on a fait varier la valeur du plus petit élément de la matrice, de 0 à 6. Tous les choix possibles sont couverts, puisque $\mathbf{p_{total}}$ = 1.

c)- il reste à interpréter ces probabilités. Trois possibilités se présentent : celle qui correspond à la probabilité unilatérale [dite à gauche], celle qui correspond à la probabilité bilatérale et enfin, celle de droite.
- la probabilité unilatérale est celle où l'effectif minimal de la matrice 2x2 varie de sa valeur à 0.
- la probabilité bilatérale regroupe la p précédente à laquelle s'ajoute la ou les valeurs, situées à droite, qui sont immédiatement inférieures à la valeur maximale de la probabilité unilatérale.
- enfin, la probabilité à droite regroupe les autres valeurs.

Dans le cas présent, le résultat du test exact de Fisher est le suivant :

Fisher's exact test
table = [20,0,5,6]
one tailed p 0.00062748
two tailed p 0.00062748
other 1
theor. value of chi2 < 3 2.13 for [0]

Le test est significatif. Nous avons pris toutefois un cas particulier, qui n'est pas le plus simple, où la valeur minimale de la matrice est déjà 0. En ce cas, bien entendu, la situation bilatérale est équivalente à la situation unilatérale et le reste de la probabilité est égale à 1. Nous allons prendre un second exemple qui permettra d'illustrer le cas général :

	A	**B**
+	20	3
-	5	6

Possible matrices and their Ps
p for 0 : [23,**0**,2,9]
1.0486E-06
p for 1 : [22,**1**,3,8]
7.2353E-05
p for 2 : [21,**2**,4,7]
0.00159176
p for 3 : [20,**3**,5,6]

0.01559928
p for 4 : [19,**4**,6,5]
0.07799642
p for 5 : [18,**5**,7,4]
0.21170456
p for 6 : [17,**6**,8,3]
0.31755684
p for 7 : [16,**7**,9,2]
0.25706982
p for 8 : [15,**8**,10,1]
0.10282793
p for 9 : [14,**9**,11,0]
0.01557999

Interprétation :

- prob. unilatérale : il s'agit de la somme des probabilités situées « entre » la valeur minimale 3 et 0. C'est donc **P_{uni}** = 0.01726445 [S p for 0 ---> p for 3].
- prob. bilatérale : il s'agit de **P_{uni}** + p for 9 [0.01557999] qui est la valeur de p immédiatement inférieure au max. de **P_{uni}**, qui est p for 3 [0.01559928]. On a donc : **P_{bil}** = 0.03284444.
- la probabilité « à droite ». Il s'agit de la somme, pour les valeurs variables de l'effectif minimal de la matrice 2x2, des probabilités qui leur sont associées :
ici : **p for 1 + p for 2 + p for 3** = 1- (1.0486E-06 + 7.2353E-05 + 0.00159176) = 0.99833483.

Voici le résultat tel qu'il apparaît dans la feuille de calcul d'Excel :

Fisher's exact test
table = [20,3,5,6]
one tailed p 0.01726445
two tailed p 0.03284444
other 0.99833483
theor. value of chi2 < 3 2.91 for [3]

Voici un dernier exemple, que nous empruntons au site : http://mathworld.wolfram.com/topics/StatisticalTests.html.

« *For an example application of the 2 x 2 test, let X be a journal, say either Mathematics Magazine or Science, and let Y be the number of articles on the topics of mathematics and biology appearing in a given issue of one of these journals. If Mathematics Magazine has five articles on math and one on biology, and Science has none on math and four on biology, then the relevant matrix would be*

	Math. Mag.	Science	
math	5	0	$R_1 = 5$
biology	1	4	$R_2 = 5$
	$C_1 = 6$	$C_2 = 4$	$N = 10.$

The sum of P-values less than or equal to $\mathbf{P}_{cutoff}$ *= 0.0238 is then 0.0476 which, because it is less than 0.05, is significant. Therefore, in this case, there would be a statistically significant association between the journal and type of article appearing.* »
[*Author: Eric W. Weisstein © 1999 CRC Press LLC, © 1999-2003 Wolfram Research, Inc.*].

Ce qui donne, avec Excel :

Possible matrices and their Ps
p for 0 : [5,**0**,1,4]
0.02380952
p for 1 : [4,**1**,2,3]
0.23809524
p for 2 : [3,**2**,3,2]
0.47619048
p for 3 : [2,**3**,4,1]
0.23809524
p for 4 : [1,**4**,5,0]
0.02380952
Fisher's exact test
table = [5,0,1,4]
one tailed p 0.02380952
two tailed p 0.04761905
other 1
theor. value of chi2 < 3 2 for [0]

De nombreux liens existent sur le test exact de Fisher ; nous avons sélectionné ceux-ci qui se distinguent par leur clarté ou la présence d'applets java :

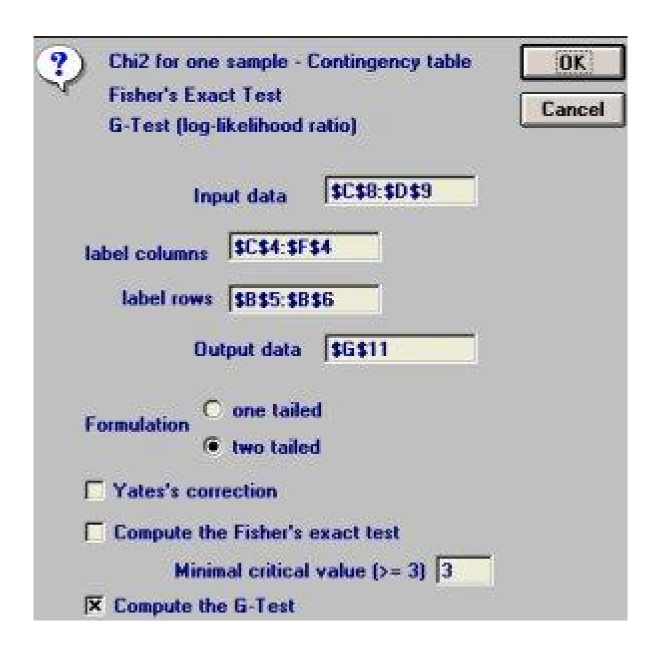
-http://www.matforsk.no/ola/fisher.htm [remarquable applet ; donne les 3 probabilités mais pas le détail du calcul des différentes probabilités] ;
- http://quantrm2.psy.ohio-state.edu/kris/nonpar.htm, **Kristopher J. Preacher and Nancy E. Briggs (May, 2001)** ***The Ohio State University*** [idem pour l'applet] ;
- http://faculty.vassar.edu/lowry/webtext.html, **©Richard Lowry 1998-2000**
- http://mathworld.wolfram.com/topics/StatisticalTests.html,*Author: Eric W. Weisstein © 1999 CRC Press LLC, © 1999-2003 Wolfram Research, Inc.*
- http://oms.b3e.jussieu.fr/biostaTGV/index.html [notez que l'applet ne marche que si vous êtes connectés, alors qu'elle fonctionne avec les deux autres, citées supra]

A noter :

-on ne peut calculer la correction de Yates ou un test exact de Fisher que pour un tableau **2** x **2**. ;

- dès qu'un effectif théorique est inférieur à 5, pour un tableau de plus de **2** x **2** cases, il faut procéder à des recoupements ou utiliser un autre test, comme celui de Kolmogorov Smirnov.
- dans le tableau de contingence, on ne doit pas avoir plus de 20% de cases où l'effectif théorique est inférieur à 5 [sinon, on procède à des regroupements ou l'on utilise le test de Kolmogorov Smirnov] et aucune case où l'effectif théorique est égal à 1.

Mode d'emploi : il suffit de sélectionner l'outil ***contingency table***. La boîte de dialogue suivante apparaît :



Vous devez saisir obligatoirement une zone d'entrée [input data] ayant une taille minimale de **2** x **2** cases et la cellule de sortie [output data]. Si vous souhaitez « *forcer* » le test exact de Fisher, il suffit d'activer la case à cocher correspondante. Vous pouvez aussi modifier la valeur théorique minimale en deçà de laquelle le test exact de Fisher est automatiquement calculé [vous ne pouvez aller au-dessous de 3]. Le test de Fisher sera calculé automatiquement si l'une des valeurs théoriques est en-deçà de la valeur minimale choisie.

2bis - Le test G ou Log - Likelihood ratio Test è ne figure pas dans l'*Analysis Tool Pack* d'Excel.

Il s'agit d'une alternative intéressante au χ^2 habituel, en ceci qu'il est plus « robuste ». En effet, dès que l'on dépasse 3 à 4 modalités différentes pour une variable, le

χ^2 est souvent significatif et l'on tire, dès lors, des conclusions hâtives, voire fautives. Il a été difficile de trouver de la bibliographie en français sur ce test. On trouve en anglais, dans le glossaire du site http://www.openlink.org/dorak/ [Common concepts in Statistics, M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.] ces lignes sur le **G-Test** :

Chi-squared test: The most commonly used test for frequency data and goodness-of-fit. In theory, it is nonparametric but because it has no parametric equivalent it is not classified as such. It is not an exact test and with the current level of computing facilities, there is not much excuse not to use Fisher's exact test for 2x2 contingency table analysis instead of Chi-squared test. Also for larger contingency tables, the G-test (log-likelihood ratio test) may be a better choice.

L'attention, comme on l'a vu, est portée sur le nombre de modalités du χ^2. Voici d'autres considérations, extraites et adaptées de : Statistical Analysis in HLA and Disease Association Studies, M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

G Statistics: An application of the **log-likelihood ratio statistics** for the hypothesis of independence in an *r* x *c* contingency table. It can also be used to test goodness-of-fit. The G-test should be preferred over Chi-squared test when for any cell in the table, $|O-E| > E$. The Chi-squared distribution is usually poor for the test statistics G^2 when **N/rc** is smaller than five (preferable to the Chi-squared test in Hardy-Weinberg Equilibrium test as long as this condition is met). StatView, HyperStat and StatXact perform G statistics.

The likelihood ratio (Chi-squared) test or maximum likelihood statistics are usually known as the **G-test** or **G-statistics** [Sokal RR, Rohlf FJ. New York: W.H. Freeman & Company, 1994]. Whenever a Chi-squared test can be employed, it can be replaced by the **G-test**. In fact, the Chi-squared test is an approximation of the log-likelihood ratio which is the basis of the G-test. Pearson originally worked out this approximation because the computation of the log-likelihood was inconvenient (but it no longer is). The Pearson's statistics, $\chi^2 = \sum [(O-E)^2/E]$ is mathematically an approximation to the log-likelihood ratio or $G = 2 \sum O \, ln \, (O/E)$

The value called G approximates to the χ^2 distribution. The G value can also be expressed as

$$G = 2\,[\sum O \, ln O - \sum O \, ln E] = 4.60517\,[\sum O \log_{10} O - \sum O \log_{10} E]$$

The **G-test** as calculated above is as applicable as a test for goodness of fit using the same number of degrees of freedom as for Chi-squared test. It should be preferred when for any cell $|O-E| > E$. For the analysis of a contingency table for independence, Wilks [Wilks SS. The likelihood test of independence in contingency tables. **Annals of Mathematical Statistics** 1935; **6**: 190-196.] formulated the calculation of the G statistics as follows:

$$G = 2\,[\sum \sum f_{ij} \, ln \, f_{ij} - \sum R_i \, ln \, R_i - \sum C_j \, ln \, C_j + N \, ln \, N\,]$$

where $\mathbf{f_{ij}}$ represents entries in each cell, $\mathbf{R_i}$ represents each row total, $\mathbf{C_j}$ represents each column total, and **N** is the sample size. The same formula can be written using logarithm base 10 as follows:

$$G = 4.60517 \left[\sum \sum f_{ij} \log_{10} f_{ij} - \sum R_i \log_{10} R_i - \sum C_j \log_{10} C_j + N \log_{10} N \right]$$

The G value approximates to χ^2 with **d.f. = (r-1)(c-1)**. When necessary, Yates' correction should still be used and the formula needs to be modified accordingly. With the exception of the above mentioned condition that ½ **O-E**½ should be smaller than E for the Chi-squared test to be valid, there is not much difference between the two tests and they should result in the same conclusion. When they give different results, the **G-test** may be more meaningful. The **G-test** has been gaining popularity in HLA and disease association studies [Klitz W, Aldrich CL, Fildes N, Horning SJ, Begovich AB. Localization of predisposition to Hodgkin disease in the HLA class II region. **American Journal of Human Genetics54**: 497-505. 1994 ; Taylor GM, Gokhale DA, Crowther D, et al. Further investigation of the role of HLA-DPB1 in adult Hodgkin's disease (HD) suggests an influence on susceptibility to different HD subtypes. **British Journal of Cancer** 1999; **80**: 1405-1411.].

C'est, comme d'habitude, dans les pages de http://www.dsa.unipr.it/soliani/soliani.html [caput3.pdf] que nous trouverons des informations beaucoup plus détaillées sur le G-Test. L'attention des statisticiens sur ce test date de la fin des années 70 [K. Larntz : ***Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics***, Journal of the American Statistical Association vol. 73, pp.253-263 ; D. A. Williams : ***Improved likelihood ratio test for complete contigency tables***, Biometrika vol. 63, pp. 33-37)]. Ces auteurs montraient que le test **G** était plus « robuste » que le χ^2 de Pearson. Sa formule générale est donnée, dans un tableau 2 x 2 [pouvant être étendue au tableau de contingence classique] par :

	Rep. X	**Rep. Y**	
Sample. Y	a	b	**n1**
Sample. y	c	d	**n2**
	n3	**n4**	**N**

$$G = 2\left[(a\ln a + b\ln b + c\ln c + d\ln d) - (n_1\ln n_1 + n_2\ln n_2 + n_3\ln n_3 + n_4\ln n_4) + (N\ln N)\right]$$

où l'on remarque 3 termes : la 1ère correspond aux entrées du tableau **2** x **2** ; la 2ème, aux sommes des colonnes et des lignes ; la 3ème, au total. Dans le cas d'un seul échantillon - équivalent au χ^2 d'ajustement - la formule devient :

$$G = 2\sum f_i \ln \frac{f_i}{\hat{f}_i}$$

où **f_i** représente la valeur observée et • $_i$ la valeur théorique, cf. équation 1. Comme pour la correction de continuité de Yates, on a écrit une correction pour le G-Test ; on l'appelle la correction de Williams et elle s'écrit :

$$q = 1 + \frac{k+1}{6N}$$

dans le cas d'un échantillon [k est le nombre de modalités]. On l'emploie quand **N** < 200.
Exemple : soit l'échantillon suivant

55 28 37 43

G Test (Log-Likelihood ratio)
N 163
k 4
df 3
G 9.403
corr. Williams 1.005
p 0.02439

Dans le cas de plusieurs échantillons - tableau de contingence - la correction de continuité de Williams s'écrit :

$$q = 1 + \frac{(m+1)\cdot(n+1)}{6N}$$

où **m** et **n** représentent le nombre de lignes et de colonnes de la matrice. [N est le nombre total de sujets].
Exemple :

« *A biologist collects leaf litter from a 1 m² quadrant placed randomly at night on the ground in each of two woodlands – 1 on **clay soil** and the other on **chalk soil**. She sorts through the leaf litter and collects woodlice belonging to 2 species, **Oniscus** and **Armadilidium**. It is assumed that the woodlice undertake their nocturnal foraging independently.* » [extrait de : ***Multinomial GOF Tests***, Roland Hall, Univeristy of Waterloo, ***Lecture 15, CDA II***]

	Oniscus	*Armadilidium*	Total
Clay soil	14	6	20
Chalk soil	22	46	68
Total	36	52	88

Traitons ces données, successivement par un χ^2 conventionnel, puis par le G-Test et enfin par un test exact de Fisher :

		nij	
	2.86	4.14	7
	0.84	1.22	2.06
nji		3.7	5.36

Contingency table

ddl 1
chi2 9.061
p 0.003

G Test (Log-Likelihood ratio)

N 88
k 2
df 1
G 8.871
corr. Williams 1.017
p 0.0029

Possible matrices and their Ps

p for 0 : [**0**,20,52,16]
2.3772E-10
p for 1 : [**1**,19,51,17]
1.4543E-08
p for 2 : [**2**,18,50,18]
3.9145E-07
p for 3 : [**3**,17,49,19]
6.1808E-06
p for 4 : [**4**,16,48,20]
6.4357E-05
p for 5 : [**5**,15,47,21]
0.00047073
p for 6 : [**6**,14,46,22]
0.00251412
p for 7 : [**7**,13,45,23]
0.01005647
p for 8 : [**8**,12,44,24]
0.03064081
p for 9 : [**9**,11,43,25]
0.07190377
p for 10 : [**10**,10,42,26]
0.13080956
p for 11 : [**11**,9,41,27]
0.18498321
p for 12 : [**12**,8,40,28]
0.20315121
p for 13 : [**13**,7,39,29]
0.17243604
p for 14 : [**14**,6,38,30]
0.11208343
p for 15 : [**15**,5,37,31]
0.05495703
p for 16 : [**16**,4,36,32]
0.01985752
p for 17 : [**17**,3,35,33]
0.00509712
p for 18 : [**18**,2,34,34]
0.00087451
p for 19 : [**19**,1,33,35]
8.9423E-05
p for 20 : [**20**,0,32,36]
4.0986E-06

Fisher's exact test

table = [6,14,46,22]
one tailed p   0.00305579
two tailed p   0.00402382
other   0.99945833
theor. value of chi2 < 3   27.82 for [6]

Il est clair que pratiquer un test de Fisher dans le cas présent est caricatural, puisque la valeur théorique inférieure pour la plus petite valeur est de 27.82 ! Il n'empèche. A des fins didactiques, nous estimons indispensable de montrer cet exemple détaillé.
Mise en oeuvre : il suffit de sélectionner dans le menu général, la case d'option correspondant à *Contingency Table*. La macro reconnaît automatiquement si l'échantillon

a 1 ou plusieurs colonnes ; de même pour le **G-Test**, toute valeur nulle [**LN**(0) = #NOMBRE!] est « sautée. »

3)- le test U de Mann et Whitney è ne figure pas dans l'*Analysis Tool Pack* d'Excel

Deux statisticiens, Mann et Whitney sont partis d'une simple constatation de bon sens : si 2 populations sont très différentes, le cas extrême se produit quand les valeurs de l'une sont inférieures aux valeurs de l'autre. Par contre, si elles sont confondues, il doit y avoir intrication des valeurs de ces 2 populations. Pour comparer deux moyennes, il faut habituellement employer le test t, qui suppose la normalité des distributions et l'égalité des variances, hypothèses invérifiables avec des effectifs faibles, comme on va le voir sur l'exemple suivant, tiré de : Armitage, P. & G. Berry (1994). ***Statistical Methods in Medical Research*** (Third Edition). Blackwell Scientific Publications, Oxford.
Dans cet exemple, on mesure le gain en poids de rats recevant des régimes différents : pour le 1[er] groupe, un régime avec des protéines de haut poids moléculaire et l'autre, un régime avec des protéines de bas poids moléculaire. L'hypothèse nulle est « *il n'y a pas de différence dans les poids moyens* ».

70	83		***Mann Whitney test***			rank x	rank y
85	97		N	19		1	2
94	104		nx	7		3	5
101	107		ny	12		4	7
107	113		Uxy	62.5		6	8.5
118	119		Uyx	22.5		8.5	10
132	123		Uo	42		11	12
	124		s2U	140		16	13
	129		mxy	101			14
	134		myx	120			15
	146		S rang xy	49.5			17
	161		S rang yx	140.5			18
			M rang xy	7.071			19
			M rang yx	11.708			
			e	1.73256622			
			critical value	18			
			critical value	12			
			p	NS			
			corrections fc	yes			
			ties	2			
			s2U	139.877193			
			e	1.73332662			
			p	0.083			

TABLEAU III
(test U de Mann et Whitney)

L'hypothèse nulle est donc conservée puisque $p > 0,05$. Le calcul consiste à ranger les observations de façon croissante et à comptabiliser les rangs, ici Uxy = 22.5. En fait, nous avons assemblé ici les résultats de deux tests

qui sont strictement équivalents : le Mann Whitney et le test de Wilcoxon [il n'a rien à voir avec le test T de Wilcoxon pour séries appariées]. Dans le tableau de droite sont regroupés les rangs [notez qu'il est inutile de ranger préalablement les observations lorsque vous saisissez le tableau des deux colonnes à traiter]. Des tables spéciales existent, qu'il faut employer dès que nx ou ny sont inférieurs à 10. Elles sont intégrées dans la feuille macro. Les valeurs limites sont indiquées pour les risques à 5% et à 1% [ici, 18 et 12].

Summary for the Mann Whitney Wilcoxon rank-sum test :

Based on ranking of all observations without regard to group associated with each observation. Can also be used with interval or ratio data that are not normally distributed
Test statistic, T, is sum of all ranks for the smaller group.

$$T = \sum_{i=1}^{n_S} R_i \quad (4)$$

where R_i is the rank of the i th observation of the smaller group and n_S is the number of observations in
the smaller group. To determine T must first rank all observations from both groups together. Tied ranks receive average of ranks that would have been spanned (e.g. if 3 observations are tied following rank 4, then each of the tied observations would receive the average of ranks 5, 6 and 7, or (5+6+7)/2 = 6; the next observation would receive rank 8). Critical values of T are based on the tails of the distribution of all possible T values (assuming no ties).
When the larger sample contains eight or more observations, distribution of T approximates a normal
distribution with mean

$$\mu_T = \frac{n_S(n_S + n_B + 1)}{2} \quad (5)$$

where n_B is the number of samples in the bigger group, and standard deviation

$$\theta_T = \sqrt{\frac{n_S n_B (n_S + n_B + 1)}{12}} \quad (6)$$

Can then construct test statistic, zT

$$z_T = \frac{T - \mu_T}{\theta_T} \quad (7)$$

which can be compared with t-distribution with infinite degrees of freedom (d.f.). This comparison is more accurate with a continuity correction where

$$z_T = \frac{|T - \mu_T| - \frac{1}{2}}{\theta_T} \quad (8)$$

En cas d'ex-aequo, la variance de (8) doit être remplacée par la formule :

$$\mathrm{var}(U) = \frac{n_s . n_b}{N(N-1)} \cdot \left(\frac{N^3 - N}{12} - \sum \frac{t^3 - t}{12} \right) \quad (9)$$

où t est le nombre de valeurs ex-aequo et $N = n_s + n_b$. A noter que cette correction n'est utile que pour de petits échantillons. L'approximation de 4 est suffisante dès que $MIN(n_s, n_b) > 8$.

4)- le test de Fisher-Yates-Terry è ne figure pas dans l'*Analysis Tool Pack* d'Excel

Ce test, selon D. Schwartz, est justifié par son efficacité. La méthode revient à normaliser les observations et à leur appliquer ensuite le test de l'écart-réduit, qui est alors le meilleur. Voyons l'exemple suivant où 2 groupes de 10 lapins, nourris chacun avec un régime enrichi en cholestérol, ont été soumis à deux traitements différents X et Y. On souhaite savoir si les résultats sont différents. Le tableau suivant montre dans la 1ère colonne les observations X, et dans la 2ème, les observations Y. Le résultat du test est affiché à droite en deux tableaux : les rangs créés à partir des nouvelles valeurs substituées aux rangs eux-mêmes. Ces rangs sont donnés par une table intégrée à la feuille macro. Au centre, les résultats, avec la valeur C1 et le carré de Z, somme indiquée en fonction du nombre total des observations. Une table existe qui donne ces valeurs, mais celles-ci sont calculées automatiquement [par exemple, pour N = 20, on trouve Z^2 = 17.7144].

						rank x	rank y
8	12		***Fisher Yates Terry***			-1.87	-1.41
13	18		N	20		-1.13	-0.75
15	19		nx	10		-0.92	-0.59
21	22		ny	10		-0.45	-0.31
23	27		Mx	21.2		-0.19	0.31
24	30		My	25.8		-0.06	0.75
25	31		C1	1.543		0.06	0.92
26	32		Z2	17.71		0.19	1.13
28	33		varC1	4.66		0.45	1.41
29	34		e	1.543		0.59	1.87
			p	0.1228			
			critical value C1 5%				
			critical value C1 1%				

TABLEAU IV
(test C1 de Fisher-Yates-Terry)

Là encore, on conserve l'hypothèse nulle puisque $p > 0,05$. On range à nouveau les observations par ordre croissant et on comptabilise les rangs. Des tables spéciales doivent être employées, intégrées dans le logiciel. Ces tables donnent les valeurs limites à partir desquelles une différence devient significative. Ces valeurs limites sont indiquées dès que nx et/ou ny sont inférieures à 10. A noter que la dénomination de test de « *Fisher Yates Terry* » est purement française et que le test n'est connu ailleurs que sous l'appellation de « *two sample rank test* ». Voici les formules grâce auxquelles le test est calculé :

$$\sigma_{C_1}^{\ 2} = \frac{n_A n_B}{N(N-1)} \sum Z^2 \quad (10)$$

L'indice utilisé est donc Z et l'on recherche la somme des Z - cf. tableau IV - soit pour les x soit pour les y. Une valeur trop petite ou trop grande indique que les x - ou les y - sont en moyenne trop grands ou trop petits. La valeur attendue de C1 est donc 0. Sa variance est donnée par l'équation précédente. Ici, on a $\mathbf{Z^2}$ = 17.71. Si l'approximation par la loi normale est permise, on a le résultat par l'écart réduit de C1 :

$$\varepsilon_{C_1} = \frac{C_1}{\sqrt{\sigma_{C_1}^{\ 2}}} \quad (11)$$

et $\mathbf{C_1}$ = 1.543 dans notre exemple. Le test C1, selon ce qu'en dit Schwartz, revient à normaliser les observations et à leur appliquer ensuite le test de l'écart-réduit.

5)- le test T de Wilcoxon, appliqué aux séries appariées è ne figure pas dans l'*Analysis Tool Pack* d'Excel

On forme pour chaque paire d'observations la différence puis on classe ces observations en valeurs absolues croissantes, en mentionnant pour chacune si elle est positive ou négative (les différences nulles sont éliminées). L'exemple suivant est tiré de : Armitage, P. & G. Berry (1994). ***Statistical Methods in Medical Research*** (Third Edition). Blackwell Scientific Publications, Oxford. L'hypothèse nulle est « *il n'y a pas de différence significative entre le médicament étudié et le placebo* ». Le résultat est affiché sur le tableau suivant :

drug	19	11	14	17	23	11	15	19	11	8
placebo	22	18	17	19	22	12	14	11	19	7

		rank +	rank -
Wilcoxon rank test			6.5
Np	10		8
Mx	14.8		6.5
My	16.1		5
P+	17	2.5	
P-	38		2.5
P0	0	2.5	
T the.	27.5	9.5	
varT	96.25		9.5
e	-1.07	2.5	
p	NS		
critical value	8		
critical value	3		

TABLEAU V
(test T de Wilcoxon)

Quand le nombre de paires est inférieure à 20 (ici, N = 10), on emploie une table spéciale, insérée dans le logiciel. A droite, la valeur du rang de chaque paire. Là encore, on doit tenir compte des ex-aequo. Nous allons donner quelques exemples supplémentaires de ce test fort utile en remployant des cas tirés de sites qui nous ont paru intéressants [à ce sujet, il est regrettable d'observer que de nombreux sites donnent les formules à employer mais point ou très peu d'exemples, ce qui nuit d'une certaine façon à la portée didactique de leur démarche]. Nous citons à présent un extrait, superposable à celui de tout à l'heure, concernant le test de Mann Whitney Wilcoxon.

Summary about the rank signed Wilcoxon test

Ordinal Data - Evaluating Two Interventions on the Same Group of Individuals : Wilcoxon Signed-Rank Test

Based on ranking of absolute differences between two observations for each individual. Test statistic, W, is sum of all ranks of differences.

$$W = \sum_{i=1}^{n} \frac{\Delta_i}{|\Delta_i|} R_i \quad (12)$$

where n is the number of individuals, D_i is the difference between observations for the i th individual,
and R_i is the rank of the absolute difference for the i th individual (note: the fraction in front of the ranks
will always have magnitude, 1, and will have the sign of the difference). If no real difference exists between individuals • observations, then the signs of the observed differences should occur by random chance; W would then compute to a number close to zero. Extreme values of W in either positive or negative sense, thus, lead to rejection of the null hypothesis that no difference exists between observations.

- Individual
- Observation One
- Observation Two
- Difference Rank of Difference
- Signed Rank of Difference

1 1600 1490 -110 5 -5
2 1850 1300 -550 6 -6
3 1300 1400 +100 4 +4
4 1500 1410 -90 3 -3
5 1400 1350 -50 2 -2
6 1010 1000 -10 1 -1

For present example W = -13 is not extreme enough to reject null hypothesis. As with other parametric methods, p-values for the Wilcoxon Signed-Rank Test are discrete in nature.

En reprenant les données des auteurs, nous trouvons le résultat suivant :

		rank +	rank -
Wilcoxon ra.	***110***	5	
Np	6	6	
Mx	1443.33		4
My	1325	3	
P+	17	2	
P-	4	1	
P0	0		
T the.	10.5		
varT	22.75		
e	1.363		
p	NS		
critical value l	0		
critical value l	0		

TABLEAU VI

(test de Wilcoxon)

For large number of individuals, however, distribution of W values approximate a normal distribution
with mean : $m_w = 0$ and variance :

$$\sigma_{P^1} = \sigma_{M^1} = \frac{1}{24} n.(n+1).(2.n+1) \quad (13)$$

From which test statistic, e_W can be computed as :

$$\varepsilon_P = \varepsilon_M = \frac{P - \frac{n.(n+1)}{4}}{\sqrt{\frac{1}{24} n.(n+1).(2.n+1)}} \quad (14)$$

which can be compared with t-distribution with infinite degrees of freedom (d.f.). Ties among the absolute differences are handled by assigning average ranks.A tie a pair creates a difference of zero. Because these are neither positive or negative, we drop such pairs from our sample.As in the case of the Wilcoxon rank sum, ties complicate funding a - value.There is no longer a usable exact
distribution for the signed rank statistic ,and the standard deviation must be adjusted for the ties before we can use the normal approximation.

a)- exemple 1 :
http://www.stat.lsu.edu/faculty/geaghan/EXST7005/Spring2002/Freelance/

On compare des concentrations d'un médicament de cellules suspendues dans des solutions différentes. Le tableau suivant montre les résultats obtenus avec notre programme :

hank's 14 14 5 7 17 27 6 15
sucrose 22 21 9 7 16 26 4 13

		rank +	rank -
Wilcoxon rank test			7
Np	7		6
Mx	13.13		5
My	14.75		
P+	10	1.5	
P-	18	1.5	
P0	1	3.5	
T the.	14	3.5	
varT	35		
e	-0.676		
p	NS		
critical value	2		
critical value	0		

TABLEAU VII
(*comparaison de deux solutions*)

Le lecteur intéressé verra que nous donnons plus d'informations que celles proposées par le rédacteur de la page. En particulier, les valeurs critiques au risque 5% et 1%.

b)- exemple 2 :
http://www.nist.gov/speech/tests/sigtests/wilcoxon.htm

Nous laissons l'auteur s'exprimer :

« *The Wilcoxon signed-rank test, like the Sign test, is a test on speakers or conversation sides or other prespecified subsets, but it takes some account the magnitude of the performance differences on each subset rather than only which system had the better performance. Table 1 offers an example of using the Wilcoxon test. For each specified subset, the difference in performance scores of systems A and B is determined, and these differences are ranked by magnitude, irrespective of sign. A mean rank is used in case of equal performance differences. A sign is then attached to each rank corresponding to the sign of the difference.* »

system A 97 88 75 90 85 94 77 89 82 90
system B 96 86 79 89 91 89 86 99 94 96

		rank +	rank -
Wilcoxon r	7	1.5	
Np	10	3	
Mx	86.7		4
My	90.5	1.5	
P+	11		6.5
P-	44	5	
P0	0		8
T the.	27.5		9
varT	96.25		10
e	-1.682		6.5
p	NS		
critical value	8		
critical value	3		

TABLEAU VIII
(*performance de deux systèmes*)

« *For large enough n (> 8 or so), Z+ has an approximately normal distribution. The test statistic is then the approximately standard normal value. Let w be the value obtained for W. The null hypothesis H0 is rejected if*

Prob (W > w) < 0.05   (one-tailed)
*Prob | (W | > | w |) = 2 * Prob (W > w) < 0.05   (two-tailed)*

*For the example above, Z+=44.0, m=27.5, s=96.25, and W=1.68. [**we find the value of e = -1.784**] At p=0.05 this is not significant as a two-tailed test, but as a one-tailed test this is significant, by a slight margin. Thus if system B is intended to be an improved version of system A, we may want to conclude that it does offer significant improvement. From tables 2 and 3 it may be observed that no such conclusion could be reached based on the sign test. We can also define Z! as the unsigned sum of the negative signed ranks. If table lookup is preferred to the normal approximation to find critical regions for the test, it is easier to use the statistic Z = min (Z+, Z!). source of this data :* G. Kanji, *100 Statistical Tests*, SAGE Publications, 1994.
*The Wilcoxon test is generally a more powerful test than the Sign test. Studies have shown that the relative efficiency of the Sign test, compared to the Wilcoxon, is 2/3 when the differences are normally distributed, and 1/3 when they are uniformly distributed (*W. Daniel, *Applied Nonparametric Statistics*, Houghton Mifflin, 1978.*, p. 36).*
This test presumes an "interval" level of measurement of performance on subsets; i.e., equal differences in performance measure between two systems mean the same, whatever the absolute level of performance. NIST has used word accuracy for this purpose, but log word error rate would better suit this requirement. It has, for example, been suggested that the difficulty of halving the error rate is approximately equal whatever the initial error rate of a system may be. »

c)-exemple 3 : tiré d'un fichier pdf. Instructif car il permet d'amplifier les résultats d'un test t « conventionnel » et d'illustrer les possibilités conjuguées de nos macros et de l'***Analysis Tool Pack*** d'Excel.

« *Researchers wished to determine if two preparations of a virus would produce different effects on tobacco*
plants. Twelve tobacco leaves were used in the experiment. Half of each leaf, selected at random, was treated
with preparation 1 of the virus. The other half of each leaf was treated with preparation 2. The response variable
measured was the number of lesions (small dark, easily counted rings) appearing on the half leaf. The numbers in the table below give the number of lesions per half leaf. »

preparation 1   55 25 21 7 16 14 7 9 5 4 4 1

preparation 2 17 17 14 11 10 9 7 5 3 2 1 0

		rank +	rank -		t-Test: Paired Two-Sample for Means		
Wilcoxon n	***38***	11				*Variable 1*	*Variable 2*
Np	11	10			Mean	14	8
Mx	14	9			Variance	220.727273	36
My	8		5.5		Observations	12	12
P+	60.5	8			Pearson Corr	0.80974555	
P-	5.5	7			Pooled Varia	72.1818182	
P0	1				Hypothesize	0	
T the.	33	5.5			df	11	
varT	126.5	2.5			t	1.96078051	
e	2.445	2.5			P(T<=t) one-t	0.03785679	
p	p < 0.05	4			t Critical one-	1.79588369	
critical value	11	1			P(T<=t) two-ta	0.07571358	
critical value	5				t Critical two-t	2.20098627	

TABLEAU IX

(*effets de préparations de virus sur des plants de tabac*)

Le test t montre une différence significative avec p # 0.037. Par contre le test **T** de Wilcoxon permet de conclure avec plus de certitude à la présence d'une différence significative.

d)- exemple 4 :
http://www.euronet.nl/users/warnar/demostatistiek/stat/kruistabellen.htm

Là encore, nous laisserons parler l'auteur :

« *-Tekentoets en Wilcoxon Signed Rank toets*

De tekentoets en de W S R toets zijn te gebruiken indien men beschikt over 2 variabelen of testgroepen met alstestvoorwaarden: gemeten op zelfde meetschaal en in dezelfde meeteenheid, tenminste ordinaal en waarvan de scores gematcht, gepaard of gekoppeld zijn. Telkens wordt van 1 respondent onder verschillende omstandigheden een meting verricht en het verschil tussen deze metingen bepaald. Het parametrisch alternatief voor deze testen is de Student TD test (Paired Two Sample T test).

Testprincipes:

Hieronder volgt een eenvoudig rekenvoorbeeld voor beide testen op basis van gegevens uit bestand: polsfrequentie (zie hieronder) (twee polsfrequentie A = voormeting, B = nameting). Het principe is om uit de 2 getallenreeksen één verschilreeks te bepalen op de score-paren (A-B). Verschillen met waarde = 0 tellen bij deze toetsen niet mee. Bij de Tekentoets gebruikt men bij de berekening alleen de + tekens. Bij de Wilcoxon Signed Rank test (WSR) wordt uit de verschilreeks een rangnummer vastgesteld voor elke score-paar en wordt een + of rangteken toegevoegd al naar gelang het verschil (A-B) groter of kleiner dan 0 is. Bij de WSR toets tellen de positieve en negatieve tekens mee. Zowel van de Tekentoets (T) als van de

Wilcoxon Signed rank toets (W) kan op de verkregen toetswaarde een Z-waarde bepaald worden met de normale benadering. »

Voormeting X 82 63 85 77 83 86 74 79 58 88
Nameting Y 78 63 82 71 73 81 74 80 58 86

		rank +	rank -
Wilcoxon r	*1*	4	
Np	7		
Mx	77.5	3	
My	74.6	6	
P+	27	7	
P-	1	5	
P0	3		
T the.	14		1
varT	35		
e	2.197	2	
p	$p < 0.05$		
critical value	2		
critical value	0		

TABLEAU X

Nous concluerons à une différence significative puisque le nombre de paires P- est inférieur à la valeur critique de 2 [**P-** = 1].

5) bis - le test du signe è ne figure pas dans l'*Analysis Tool Pack* d'Excel

Il est à mettre au même rang - si l'on peut dire - que le test T de Wilcoxon dont il constitue en somme une généralisation. Là encore, nous mettrons à contribution l'excellent aide-mémoire que nous avons déjà signalé lors de l'examen du test des séquences de Wald Wolfowitz.
Le test du signe consiste à tester l'égalité de la médiane m à une valeur $\mathbf{m_0}$ connue, avec un échantillon de variables $\mathbf{X_i}$, dont la médiane **m** est inconnue. On trouvera infra le test de la médiane [*Mood Test*] pour deux échantillons, généralisable à plusieurs échantillons [*extensed median test*].

Trois hypothèses $\mathbf{H_0}$ sont possibles selon que :
- $m = m_0$ contre $\mathbf{H_1}$: $m \neq m_0$ [alternative bilatérale] ;
- $m \leq m_0$ contre $\mathbf{H'_1}$: $m > m_0$ [alternative unilatérale] ;
- $m \geq m_0$ contre $\mathbf{H''_1}$: $m < m_0$ [alternative unilatérale].

Ce test ne peut pas s'utiliser avec de petits échantillons [lorsque $n < 6$] car l'on ne peut alors dégager de région de rejet dont la probabilité soit inférieure à 5%.
Le test consiste à faire la somme des observations supérieures à $\mathbf{m_0}$, notée alors **S+** et des autres

observations, notées **S-**. Sous l'hypothèse nulle H_0, il est intuitif que **S+** et **S-** soient proches de **n/2**. Dès lors, la méthode consiste à observer à partir de quelle valeur, forcément proche de 0, on met en évidence pour **S+** et/ou **S-** une différence significative. Les valeurs critiques de **S+** et **S+,S-** sont tabulées sur la table de probabilités conditionnelles cumulées jusqu'aux valeurs : S_{min} = 10 et **n** = 20. A partir de **n** = 20, on peut employer la formule suivante, donnant **Z** :

$$Z = \frac{S_{min} + \frac{1}{2} - \frac{n}{2}}{\sqrt{n}/2} \quad (15)$$

Voici un exemple qui tire parti à la fois du test T de Wilcoxon [rang et signe] et du test du signe ; il est extrait de : ***Biostatistics for Fellows***, 2001, *Overview of Nonparametric Statistical Tests 2.0* :

« *Residents are rated on their ability to insert CVCs in the ICU prior to and after an intensive one month ICU rotation with a critical team. The scale ranges from 1 (dangerous) to 10 (highly proficient).* »

ranking before 3 5 6 1 2 4 3 7 5 4 4
ranking after 6 7 6 6 8 5 9 6 6 7 7

Wilcoxon rank test

Np 10
Mx 4
My 6.64
P+ 2
P- 53
P0 1
T the. 27.5
varT 96.25
e -2.599
p $p < 0.01$
ties 8
critical value Pmin 5% 8
critical value Pmin 1% 3

Sign Test

S+ 1
S- 9
S min 1
S max 9
H1bil. 0.011
H'1 uni 0.011
H"1 0.999

6)- le test de Kruskal et Wallis è ne figure pas dans l'*Analysis Tool Pack* d'Excel

Ce test, très utile, permet d'analyser la liaison entre un caractère quantitatif et un caractère qualitatif à **k** classes (k >2).

« *The original paper is Kruskal and Wallis (1952) with errata appearing in Kruskal and Wallis (1953). No attempt is made to calculate exact values, rather an incomplete beta approximation is used following Wallace (1959).* » [Bob Wheeler, *The SuppDists Package* February 19, 2002]

Ce test permet notamment d'effectuer des comparaisons multiples en testant ce que l'on appelle les rangs moyens. C'est ici le lieu, néanmoins, de mettre en garde les utilisateurs contre le danger de la « pêche à la ligne » des comparaisons multiples : il y a en effet, un grand danger à comparer sans discernement, plusieurs échantillons. Nous dirons que ces comparaisons multiples font appel à ce que l'on appelle la notion de contraste. L'exemple suivant est tiré de : Zar, J. H. (1984). ***Biostatistical Analysis*** (Second Edition). Prentice Hall. Un chercheur veut savoir si l'hypothèse nulle : « *l'abondance des mouches est semblable dans les trois couches de végétation* », à partir des 3 variables suivantes : herbes - arbrisseaux et arbres. Si l'hypothèse nulle est rejetée, le chercheur veut savoir aussi quelle est (quelles sont) la (les) variable(s) qui est (sont) différente(s) de(s) l'autre (autres). On obtient le tableau suivant, à partir de 5 observations pour chaque variable :

1								
2	A	B	C		A	B	C	
3	14	8.4	6.9		15	11	8	
4	12.1	5.1	7.3		14	2	9	
5	9.6	5.5	5.8		12	4	5	
6	8.2	6.6	4.1		10	7	1	
7	10.2	6.3	5.4		13	6	3	
8					A	B	C	
9				sample size	15			
10				count	5	5	5	
11				average	10.82	6.38	5.9	
12				rank sum	64	30	26	
13				rank average	12.8	6	5.2	
14				ties	0			
15				Q	8.72			
16				Q corr.	8.72			
17				p	0.0128			
18								
19						A	B	C
20					B	++ 6.8 | 6.77		
21					C	++ 7.6 | 6.77	– 0.8 | 6.77	
22								

TABLEAU XI
(test de kruskal-Wallis)

On observe que les 3 séries d'observation sont différentes dans l'ensemble avec p = 0.013. Cela autorise, à titre

indicatif seulement, à voir quelles sont les séries qui diffèrent l'une de l'autre : ici, on trouve que la série 1 diffère de la série 2 et que la série 1 diffère de la série 3 (++ sur le tableau du bas) ; en revanche, la série 2 ne diffère pas de la série 3 (-- sur le tableau du bas). Les rangs de chaque colonne sont indiqués dans le tableau de droite ; au-dessous, moyennes, somme des rangs, etc. Prise en compte des ex-aequo. Dans le tableau du bas, valeurs des moyennes comparées et valeur de la limite à partir de laquelle la différence est significative [ici, 6.77]. Voici la marche à suivre pour les calculs :

- on range les observations comme dans le test de Mann-Whitney Wilcoxon [ne pas confondre avec le signed rang T of Wilcoxon] en tenant compte des ex aequo.
- on calcule les scores moyens pour chaque colonne :

$$\overline{R}_j = \frac{\sum_{i=1}^{n_j} R_{ji}}{n_j} \quad (16)$$

où Rji est le rang de l'observation i du groupe j, nj en étant l'effectif.
- on compare **Rj** avec la somme théorique $\mathbf{R_T}$:

$$\overline{R}_T = \frac{\sum_{i=1}^{N} i}{N} = \frac{N+1}{2} \quad (17)$$

où N est le nombre total d'observations. On nomme **D** la somme de carrés pondérée :

$$D = \sum_{j=1}^{m} n_j \left(\overline{R}_j - \overline{R}_T\right)^2 \quad (18)$$

- on détermine la statistique **H** [ou **Q**] qui suit approximativement une loi du chi2 à **m-1** ddl, où m est le nombre de groupes [i.e. de colonnes].

$$H = \frac{D}{N(N+1)/12} = \frac{12}{N(N+1)} \sum_{j=1}^{m} n_j \left(\overline{R}_j - \overline{R}_T\right)^2 \quad (19)$$

Cette équation s'entend sans ex aequo ; s'il y en a, la variance doit être augmentée d'une certaine quantité :

$$C = 1 - \frac{\sum t^3 - \sum t}{N(N^2 - 1)} \quad (20)$$

où t représente le nombre d'ex aequo. **H** corrigée s'obtient alors par : **H' = H/C**.

- si H est significatif [$p < 0.05$], on peut, à titre indicatif, « aller voir » quelle moyenne est différente de l'autre [ou des autres]. Pour cela, on pratique un test qui s'inspire de la méthode des contrastes et qui consiste à comparer la différence observée entre les moyennes à une certaine quantité :

$$|\bar{r}_i - \bar{r}_k| > z_{\alpha'} \left[\left(\frac{1}{n_i} + \frac{1}{n_k} \right) . \frac{n(n+1)}{12} \right]^{0.5} , \alpha' = \frac{\alpha}{I.(I-1)} \quad (21)$$

où I est le nombre de colonnes.

Conditions d'application de la statistique de kruskal Wallis : **nj >= 5** pour **m = 3** et **N >= 10** pour m = 4. Des tables spéciales doivent être consultées pour des valeurs inférieures. Ces tables sont intégrées à la feuille macro. Elles peuvent être consultées à l'adresse suivante : http://cons-dev.univ-lyon1.fr/Enseignement/Stat/Tables/TabKrusk.html.ou à l'URL suivante : http://www.psychology.nottingham.ac.uk/courses/modules/statsguides/StatisticalTables.htm.

Voici un exemple :

TRT A	8	9	11	45	1
TRT B	1	2	0	5	1
TRT C	0	-2	-3	5	

Ces trois traitements donnent-ils des effets différents, appliqués à 5 malades dans le groupe A, à 5 autres dans le groupe B et à 4 autres dans le groupe C ?

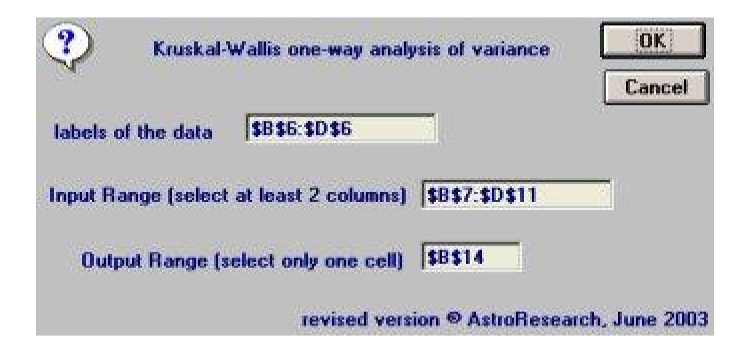
<table>
<tr><td></td><td>A</td><td>B</td><td>C</td><td></td></tr>
<tr><td></td><td>13</td><td>6</td><td>3.5</td><td></td></tr>
<tr><td></td><td>10.5</td><td>8</td><td>2</td><td></td></tr>
<tr><td></td><td>14</td><td>3.5</td><td>1</td><td></td></tr>
<tr><td></td><td>10.5</td><td>10.5</td><td>10.5</td><td></td></tr>
<tr><td></td><td>6</td><td>6</td><td></td><td></td></tr>
<tr><td></td><td>A</td><td>B</td><td>C</td><td></td></tr>
<tr><td>sample size</td><td>14</td><td></td><td></td><td></td></tr>
<tr><td>count</td><td>5</td><td>5</td><td>4</td><td></td></tr>
<tr><td>average</td><td>5</td><td>2</td><td>0</td><td></td></tr>
<tr><td>rank sum</td><td>54</td><td>34</td><td>17</td><td></td></tr>
<tr><td>rank average</td><td>10.8</td><td>6.8</td><td>4.25</td><td></td></tr>
<tr><td>ties</td><td>9</td><td></td><td></td><td></td></tr>
<tr><td>Q</td><td>5.666</td><td></td><td></td><td></td></tr>
<tr><td>Q corr.</td><td colspan="3">5.859 | 5% > 5.643 | 1% > 7.791</td><td></td></tr>
<tr><td>p</td><td>p < 0.05</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td></td><td>B</td><td>-- 4 | 6.33</td><td></td><td></td></tr>
<tr><td></td><td>C</td><td>-- 6.55 | 6.72</td><td>-- 2.55 | 6.72</td><td></td></tr>
</table>

TABLEAU XII

(test de kruskal Wallis)

Ce tableau donne à voir le classement en rangs, et plus bas, les résultats. On voit que la valeur de **Q = 5.859** [corrigée pour les ex aequo] et significative à 5% puisque la valeur correspondante dans la table est de **5.643** au risque 5%. Elle est par contre inférieure à celle relative au risque 1% [7.791]. Toutefois, le manque de puissance ne permet pas, lorsqu'on recherche une différence entre rangs moyens, de relever une quelconque différence.

Mise en oeuvre du test :



Il suffit de remplir les 3 blocs de saisie [les étiquettes sont à saisir dans la première ligne].

7)- le test de Wald-Wolfowitz è ne figure pas dans l'*Analysis Tool Pack* d'Excel

Ce test sert à contrôler une séquence. L'exemple suivant est emprunté à : Cohen, L. & M. Holliday (1983). ***Statistics for Social Scientists***. Harper & Row. Voici un exemple pour

deux échantillons tirés de l'ouvrage remarquable : ***Introduction à la décision statistique - Fiches aide - mémoire***, ouvrage collectif, 1997. Il s'agit en fait de bien plus que des fiches aide-mémoire ! L'ouvrage peut être consulté sur le suite suivant : http://www.inra.fr/bia/ftp/T/FPstat/module1/version2/Fiches.ps.gz.
Laissons parles un peu les auteurs :

« *On range ensemble les deux échantillons par ordre croissant, et on remplace chaque valeur par le numéro de son échantillon d'origine. On obtient ainsi une suite de m + n symboles 1 ou 2. On définit un run comme une séquence d'observations de l'un des deux échantillons (une séquence de symboles 1 ou de 2) immédiatement suivie soit d'une séquence d'observations de l'autre échantillon, soit de rien (fin de la suite des m + n symboles). La statistique utilisée est le nombre total R de runs dans la suite construite à partir des deux échantillons. Sous Ho, les Xi et les Yj sont tous issus d'une même population et on ne doit pas s'attendre à observer un nombre de runs trop faible, contrairement à certaines alternatives (cas limites : seulement 2 runs si la différence des positions est très forte par rapport aux dispersions • par exemple 111111-222222 ; ou 3 runs si la position est commune mais si les dispersions sont très différentes • par exemple 111-222222-111').* »

Par exemple, si nous prenons les deux suites :

X 1 4 5 8 9
Y 2 3 4 6 7 10

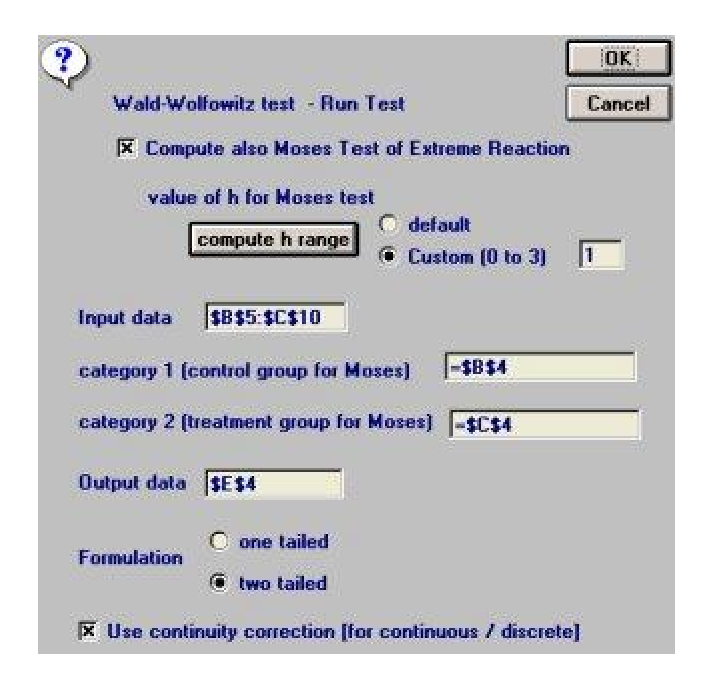
nous obtenons le résultat suivant

Value	Category	Rank	Sequence	Runs		
1	X	1	1	1	***Wald Wolfowitz test***	
2	Y	2	0	0	N	11
3	Y	3	0	1	n1	5
4	X	4.5	1	1	n2	6
4	Y	4.5	0	1	runs	8
5	X	6	1	1	µ	6.455
6	Y	7	0	0	s	1.559
7	Y	8	0	1	Z	1.312
8	X	9	1	0	p	0.18952
9	X	10	1	1	exact value	0.911
10	Y	11	0			

TABLEAU XIII
(test de Wald-Wolfowitz sur deux séries)

Le résultat montre qu'il n'y a pas de différence significative ($p = 0,251$).

Mise en oeuvre : il faut appeler la boîte de dialogue suivante :



Input data : saisie d'une ou de deux colonnes, selon que les éléments sont déjà disposés sous une forme ordinale. Dans le cas présent, la sélection correspond aux deux colonnes contiguës X et Y.
category 1 et 2 : on peut saisir n'importe quel caractère ou une référence. Ici, $B$4 correspond à X et $C$4 à Y. La macro sait faire la différence entre la saisie d'un texte, d'un nombre ou d'une référence.
output data : zone de réponse : le tableau XIII permet d'observer que les résultats sont disposés sur 5 colonnes pour l'analyse et de deux pour la synthèse. Dans la première colonne, la suite des nombres est réordonnée ; la colonne suivante montre la catégorie correspondante ; la colonne suivante, le rang correspondant [la même macro est employée pour tous les calculs de rang]. Les deux colonnes suivantes montrent la transposition des valeurs ordinales en 1 et 0 ; la dernière colonne, enfin, détermine le nombre de runs, ici 8. Dans le cas d'ex aequo - ici le rang moyen 4.5 qui correspond à la valeur 4, il faut privilégier la suite des symboles en sorte de rendre **R** le plus grand possible, ce qui signifie : privilégier l'hypothèse nulle $\mathbf{H_0}$.
continuity correction : permet de tenir compte du fait qu'une distribution continue est utilisée dans l'approximation d'une distribution discrète, selon que les runs soient réparties de façon « plutôt » unilatérale ou bilatérale.
Wald Wolfowitz : **U = 8** runs. La macro renvoie le résultat en valeur **Z** et la valeur exacte, tabulée.

Voyons un second exemple, tiré du même ouvrage, remployé de Sprent, 1992 : les auteurs se demandent si la distribution du nombre de pages est identique dans deux populations de livres : livres de statistiques et livres divers :

divers 29 39 60 78 82 112 125 170 192 224 263 275 276 286 369 756
stat 126 142 156 228 245 246 370 419 433 454 478 503

Wald Wolfowitz test

N 28
n1 12
n2 16
runs 7
µ 14.714
s 2.541
Z -2.839
p 0.00453
exact value not available

La valeur **Z** = 2.839 autorise à conclure que les livres de statistiques sont paginés de manière différente que les livres « divers. »
Notez que ce test est équivalent au « run test » ou test des signes [nous avons 7 runs dans le dernier exemple]. La différence entre le test de Wald Wolfowitz et le run test a trait simplement au fait que, dans le run test, les données binaires 1 et 0 sont déjà mises en place : le traitement des données requiert une seule colonne ; alors que dans le test de Wald Wolfowitz, les données sont d'abord « brutes », en deux colonnes séparées qu'il faut ordonner en 1 et 0. Le problème vient alors des ex aequo ; s'ils sont situés dans une même colonne, ils sont considérés comme d'habitude ; s'ils sont situés dans deux colonnes différentes, il faut les ordonner en sorte de privilégier l'hypothèse nulle $\mathbf{H_0}$. Cela requiert davantage de temps ; notez enfin que le test de Wald Wolfowitz est moins puissant que le test de Kolmogorov Smirnov. Les formules permettant de déterminer le Z sont les suivantes, sachant que l'approximation normale est atteinte quand $\mathbf{n_1}$ et $\mathbf{n_2}$ > 10.

$$\mu = \frac{2n_1 n_2}{n_1 + n_2} + 1 \quad (22)$$

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$ (23)

$$Z = \frac{\left(u - \mu + \frac{1}{2}\right)}{\sigma}$$ (24)

Notez encore que le test de la médiane se rapproche du test des signes. Et qu'enfin, il existe un test qui permet de relever si deux échantillons ont une même variance. Ce test est connu comme le ***Moses test for equal variability***. Il est congénère du test de Wald Wolfowitz et se rapproche, par sa mise en oeuvre du test de Mann Whitney U parce qu'il est basé sur un classement en rang qui s'inspire absolument de celui qui est visible au tableau VII. Il y a peu d'information qui circule sur le web pour ce test : on peut citer :

- un fichier pdf dont seule la version en cache est disponible ; hélas, les formules en sont illisibles. C'est un fragment de cours, débité en plusieurs leçons, qui tire son origine de : http://cui.unige.ch/~koval/taras/Param%20Nonparam%20Stat%20Proc/test15.pdf. En voici un extrait :

« Developed by Moses (1963), the Moses test for equal variability is a nonparametric procedure that can be employed in a hypothesis testing situation involving two independent samples. If the result of the Moses test for equal variability is significant, it indicates there is a significant difference between the sample variances, and as a result of the latter the researcher can conclude there is a high likelihood that the samples represent populations with different variances. The Moses test for equal variability is one of a number of tests of dispersion (also referred to as tests of scale or spread) that have been developed for contrasting the variances of two independent samples. [...]. Some sources recommend the use of nonparametric tests of dispersion for evaluating the homogeneity of variance hypothesis when there is reason to believe that the normality assumption of the appropriate parametric test for evaluating the same hypothesis is violated. Sources that are not favorably disposed toward nonparametric tests recommend the use of Hartley • s*F* test for homogeneity of variance/*F* test for two population variances »

Le test permet de tester la différence des rangs pour une variable donnée organisée en deux groupes. Le groupe « contrôle » est défini comme étant celui qui contient la valeur la plus basse ; l'autre groupe est étiqueté comme « expérimental ». Les deux groupes sont combinés exactement de la même manière que pour le test de Wald Wolfowitz. On définit alors une quantité nommée par les anglo-saxons le « *span* », que l'on peut traduire par

intervalle, quoique le terme, stricto sensu, soit impropre. Il serait sans doute plus avisé de dire « *l'écart* » ou même mieux, la « *distance* ». Cette distance - *span* - représente donc la différence, calculée en rang, entre la valeur minimale et la valeur maximale du groupe contrôle + 1, ce que l'on peut écrire :

SPAN = Rank(Largest Control Value) - Rank(Smallest Control Value) + 1

Valeur qui doit être arrondie à l'entier le plus proche [sur Excel, on utilise la fonction **INT()** ou **TRUNC()**]. Je laisse à nouveau la parole à l'auteur du site sur lequel on peut voir le document suivant :

« *Under the null hypothesis, therefore, the very long leads, the very short leads, as well as the moderate leads, should all contain a mixture of E's and C's. Under the alternative hypothesis, most of the E's will be low, or most of the E's will be high, or most of the E's will be low or high with the C's concentrated in the middle. Under these three conditions, the C's will be concentrated at the high end, the low end, and the middle region, respectively. The Moses test determines whether the C's are so closely compacted or congested relative to the nE + nC leads as to call for a rejection of the null hypothesis that both E's and C's come from the same population.*
To perform the Moses test, the leads from the E and C indicators are combined and arranged in a single ordered series, retaining the identity of each lead as having come from the C indicator or the E indicator. Then the span of the C leads is determined by noting the lowest and highest C leads and counting the number of leads between them, including both extremes. Thus the span s' is defined as the smallest number of consecutive leads in an ordered series needed to include all the C leads. Since the sampling distribution of s' is known (Moses, 1952), it may be used for tests of significance. Because s' is essentially the range of the C leads, and the range is known to be unstable, the modification suggested by Moses is to pick an arbitrary small number h, in advance of the analysis. A value such as 1 or 2 is typical. Then, the span sh of the C leads is determined after dropping the h most extreme C ranks. .. » [***THE LEAD PROFILE AND OTHER NON-PARAMETRIC TOOLS TO EVALUATE SURVEY SERIES AS LEADING INDICATORS***, Anirvan Banerji Economic Cycle Research Institute New York, 24 th CIRET Conference Wellington, New Zealand March 17-20, 1999]

On a donc : $\mathbf{g = S_h - (n_c - 2h)}$ où $\mathbf{S_h}$ représente la distance [span], $\mathbf{n_c}$ l'effectif du groupe contrôle et **h** ce que j'appellerai, faute de mieux le « *facteur de Moses*.» La valeur de **h** peut être attribuée par défaut à $\mathbf{0.05 \times n_c}$ en prenant la partie entière; elle peut aussi être comprise entre 1 et 3. On calcule ensuite les probabilités conditionnelles cumulées grâce à la formule suivante :

$$p(s_h \leq n_C - 2h + g) = \frac{\sum_{i=0}^{g} \binom{i + n_C - 2h - 2}{i} \binom{n_E + 2h + 1 - i}{n_E - i}}{\binom{n_C + n_E}{n_C}}$$

(25)

On calcule **p** pour **h = 0** puis :

« *The test is repeated, dropping the **h** lowest and **h** highest ranks from the control group. If not specified by the user, **h** is taken to be the integer part of **0.05 x nc** or **1**, whichever is greater. If **h** is user specified, the integer value is used unless it is less than one. The significance level is determined as above.* » [NPAR TEST in SPSS]

Les anglo-saxons emploient le terme de « *truncated span* » selon la valeur attribuée au facteur **h**. Il s'agit en fait de l'équivalent des « *outliers* » ou données réputées aberrantes.

- <u>exemple 1</u> : soit deux séries **C** et **E** ; on veut observer si les variances de C et de E diffèrent :

C 12 16 6 13 13 3 10 10 11
E 25 5 14 19 0 17 15 8 8

Résultat :

Value 0 3 5 6 8 8 10 10 11 12 13 13 14 15 16 17 19 25
Category E C E C E E C C C C C C E E C E E E
Rank 1 2 3 4 5.5 5.5 7.5 7.5 9 10 11.5 11.5 13 14 15 16 17 18
Sequence 0 1 0 1 0 0 1 1 1 1 1 1 0 0 1 0 0 0
Runs 1 1 1 1 0 1 0 0 0 0 0 1 0 1 1 0 0

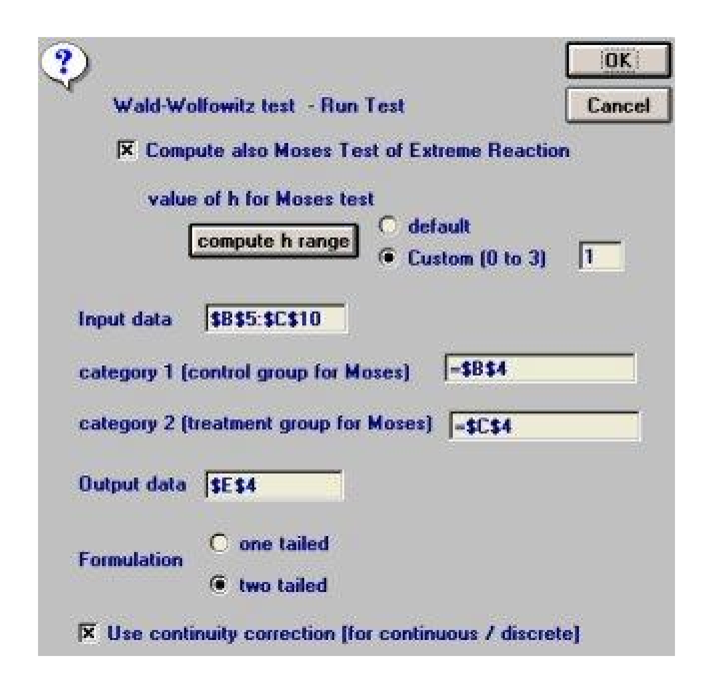
Wald Wolfowitz test
N 18
n1 9
n2 9
runs 9
µ 10
s 2.058
Z -0.243
p 0.80801
exact value 0.39

Moses Test Extreme Reaction
sh 9
nc 9

2h 2
g 2
p 0.0767174

- <u>exemple 2</u> : soit deux séries **T** et **P** où **T** figure un groupe recevant un traitement réel et **P** un autre groupe recevant un placebo ; on veut savoir si les variances diffèrent

Mise en oeuvre : on appelle la boîte de dialogue suivante :



Cette boîte est évidemment la même que pour le test de <u>Wald-Wolfowitz</u>. On y distingue une case à cocher pour prendre en compte le ***Moses Test*** ; puis le facteur **h** à saisir comme vu plus haut, avec une possibilité permettant de faire varier à volonté la probabilité pour un h différent [bouton compute h range for censoring of range]. Le reste rejoint ce que nous avons déjà dit. En principe, le programme reconnaît automatiquement le groupe contrôle qui possède la valeur la plus basse. Voici les données et les résultats :

T 64 7 73 56 82 68 94 68 29 36 81 32 39 88 38 53
P 26 23 53 74 -9 23 30 96 -6 31 47 63 21 90 26 -4 -2 49 70 98 92 7 37

Value -9 -6 -4 -2 7 7 21 23 23 26 26 29 30 31 32 36 37 38 39 47 49 53 53 56 63 64 68 68 70 73 74 81 82 88 90 92 94 96 98
Category P P P P T P P P P P P T P P T T P T T P P T P T P T T T P T P T T T P P T P P

Rank 1 2 3 4 5.5 5.5 7 8.5 8.5 10.5 10.5 12 13 14 15 16 17 18 19 20 21 22.5 22.5 24 25 26 27.5 27.5 29 30 31 32 33 34 35 36 37 38 39

Sequence 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 1 0 1 1 0 0 1 0 1 0 1 1 1 0 1 0 1 1 1 0 0 1 0 0

Runs 0 0 0 1 1 0 0 0 0 0 1 1 0 1 0 1 1 0 1 0 1 1 1 1 1 0 0 1 1 1 1 0 0 1 0 1 1 0

Wald Wolfowitz test

N 39
n1 16
n2 23
runs 21
µ 19.872
s 2.979
Z 0.546
p 0.58507
exact value not available

Moses Test Extreme Reaction

sh 21.5
nc 23
2h 2
g 0
p 2.5695E-0

8)- le test de corrélation de rang de Spearman è ne figure pas dans l'*Analysis Tool Pack* d'Excel

Ce test est l'équivalent du coefficient de corrélation pour des effectifs de taille inférieur à 30. Les données de base consistent en une série de n couples (xi, yi). Il convient ici de classer séparément les x et les y : à chaque x correspond alors un rang allant de 1 à n ; idem pour chaque y. Là encore, on départage les ex-aequo selon un procédé analogue à celui des autres tests [non compris le test T de Wilcoxon pour séries appariées]. Voici un exemple, emprunté à Cohen, L. and Holliday, M. ***Statistics for the Social Scientists*** [Harper & Row, London, 1982, p. 155].

4								
5					rank x	rank y	LS for ranked data	
6	30	70	*Spearman rank order c*		2	2.5	2.26	
7	30	70	N	24	2	2.5	2.26	
8	25	68	Np	12	2	2.5	2.26	
9	27	63	Mx	26	4	2.5	4.14	
10	23	52	My	63	5.5	8	5.56	
11	21	50	Sd	0	5.5	5.5	5.56	
12	27	68	Sd2	15	7.5	5.5	7.44	
13	23	59	df	10	7.5	7	7.44	
14	23	52	r'	0.948	10	10.5	9.8	
15	30	70	t	9.419	10	9	9.8	
16	28	70	p	0	10	10.5	9.8	
17	25	64	Z	0.00084	12	12	11.68	
18			ties	18				
19			r' corr.	0.945				
20			t corr.	9.179				
21			p corr.	0				

TABLEAU XIV
(*coefficient de corrélation des rangs, ou de Spearman*)

A gauche, en mode de sélection, les données apparaissent. Elles doivent figurer dans deux colonnes contiguës. A droite, les résultats. On dispose de :

N : nombre d'observations - **Np** : nombre de paires - **Mx** : moyenne des x - **My** : moyenne des y - **Sd** : somme des différences de chaque rang - **sd^2** : somme des carrés de chaque rang - **df** : nombre de degrés de libertés - **r'** : coefficient de corrélation de Spearman - **t** : statistique de Student - **p** : probabilité associée - **Z** : statistique de la loi normale - **ties** : nombre d'ex-aequo et corrections pour les ex-aequo : r', t et p.

Ensuite les colonnes des rangs puis la droite de régression selon : **y = ax + b**. On dispose aussi d'un graphique :

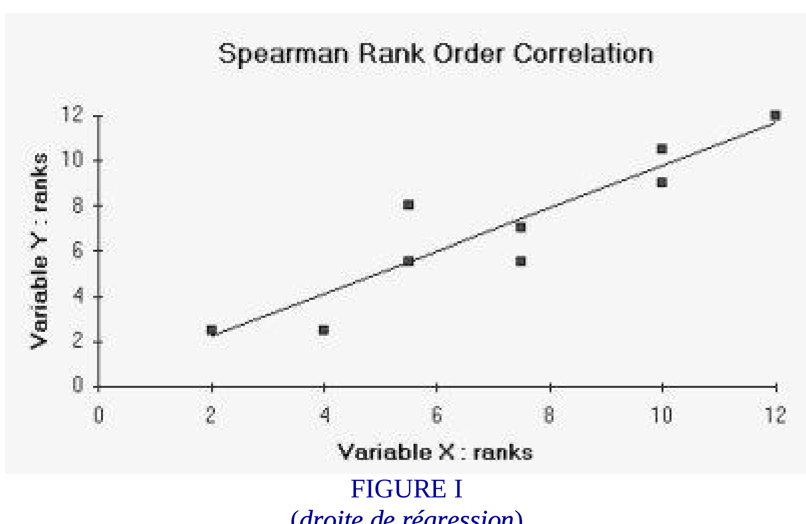


FIGURE I
(*droite de régression*)

D'autres tests doivent être prochainement disponibles ainsi que des liens faisant le point de ce que l'on peut trouver sur le sujet. Il s'agit de la première statistique, calculée à partir des rangs, qui ait été proposée (1903). Soit n le nombre d'items à classer; $\mathbf{X_i}$ est le rang de l'item **i** dans un premier classement et $\mathbf{Y_i}$ son rang dans un second. $\mathbf{Dj = |X_i - Y_i|}$ est la différence de rangs observés entre les deux classements. Le coefficient de corrélation des rangs de Spearman a pour expression :

$$R_S = 1 - 6 \, \Sigma \, D_i^2/[n(n^2 - 1)] \quad (26)$$

Plus Rs est proche de 1, plus les deux classements sont proches; A la limite, ils sont complètement identiques si $\mathbf{R_s = 1}$. Au contraire, plus $\mathbf{R_s}$ est proche de 0, plus les deux classements sont indépendants. En cas d'ex aequo, on doit calculer la quantité :

$$\sum_{i=1}^{n} (y_{ki} - \bar{y}_k) = \frac{n^2 - n}{12} - \sum_{i=1}^{n} \frac{t_{ki}^3 - t_{ki}}{12} \qquad k = 1, 2$$

(27)

où $\mathbf{tk_i}$ est le nombre d'ex aequo au rang **i** de la variable $\mathbf{X_k}$.

9)- le test de Friedman è ne figure pas dans l'*Analysis Tool Pack* d'Excel

Ce test est important à connaître car il est utilisé dans des circonstances spéciales, souvent rencontrées en pratique : les tests de dégustation ou si l'on préfère, et de manière plus large, les tests « *d'appréciation.* » Le calcul des rangs, nécessaire, à ce test, par groupes [blocs et traitements] fait en outre que ce test se rapproche de la statistique de Page ainsi que du coefficient de concordance **W** de Kendall. Voyons un premier exemple.

a)- l'hypnose : dans une expérimentation pratiquée en 1975, Lehman a enregistré le « potentiel cutané » en millivolts chez 8 sujets qui, par ailleurs, étaient interrogés sur la coloration psychique « crainte, joie, tristesse et joie » sous hypnose.
[site consulté : http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/sect41.htm].

Voici le tableau des observations :

	fear	joy	sadness	calmness
1	23.1	22.7	22.5	22.6
2	57.6	53.2	53.7	53.1
3	10.5	9.7	10.8	8.3
4	23.6	19.6	21.1	21.6
5	11.9	13.8	13.7	13.3
6	54.6	47.1	39.2	37
7	21	13.6	13.7	14.8
8	20.3	23.6	16.3	14.8

Comme on peut le voir sur le tableau XV, le résultat n'est pas significatif.

Friedman's Test

N	32
N (=t)	8
k (=b)	4
SRj2	1686
Fr (3, 21)	2.57
p	0.08129
critical value 5%	7.5
ties	no
corr. for ties	1
T'	6.45
p'	0.09166

Friedman's Test			1	2	3	4
N	32	fear	4	4	3	4
N (=t)	8	joy	3	2	2	1
k (=b)	4	sadness	1	3	4	2
SRj2	1686	calmness	2	1	1	3
Fr (3, 21)	2.57					
p	0.08129					
critical value	7.5			fear	joy	sadness
ties	no		joy	−7 | 21.64		
corr. for ties	1		sadness	−8 | 21.64	−1 | 21.64	
T'	6.45		calmness	−13 | 21.64	−6 | 21.64	−5 | 21.64
p'	0.09166					

TABLEAU XV
(*statistique de Friedman*)

Le tableau XV peut être lu en trois sous-ensembles : à gauche, les résultats proprement dits [p # 0.09, donc NS mais il peut être utile de donner la valeur au lieu du NS classique]. On peut lire **N** [nombre total : 32] ; **t** [nombre de « traitements », ici 8, encore appelé **N**] ; **b** [nombre de blocs, ici 4 : « fear - joy - sadness - calmness », encore appelé **k**]. Puis : $S\mathbf{R}_j^2$ [somme des carrés du total des **Rj** : $27^2 + 20^2 + 19^2 + 14^2$] ; le **T** de Friedman est ici inférieur à la valeur critique 7.5 [lu dans la table inclue dans la feuille macro]. A noter que la table donnée par ce site est incomplète ; nous avons trouvé les valeurs correspondant à **k** = 3 (suivant n > 8), **k** = 4 (suivant n > 4) et enfin **k** = 5 (n de 3 à 6) à la page suivante : http://www-class.unl.edu/psycrs/handcomp/hcfried.PDF.

Poursuivons : dans le cas présent, il n'y a pas d'ex-aequo ; sinon, il faut appliquer une correction qui apparaît sous le label **T'**. Pour mémoire, même lorsque le test est non significatif, nous avons donné à titre indicatif seulement les comparaisons multiples - chose interdite en statistique !... - avec la valeur critique. Notez encore que le résultat **T** de la statistique de Friedman peut être rapporté par un test **F** lorsque **k** est compris entre 3 et 5 ou que **N** est compris entre 5 et 10. Ce test **F** est lu pour (N -1 ; k -1) ddl, cf. infra, (30).

b)- extrait de Reading/Reference (for all SPSS-related matters): Kinnaer, P.R. and Gray, C.D. (1997). *SPSS for Windows made simple* (second edition). Hove: Psychology Press.

Un chercheur intéressé par l'apprentissage du langage veut savoir si l'âge d'acquisition affecte les sentiments des sujets qu'un mot est familier. Il conçoit une expérience avec les mots qui sont généralement acquis entre des âges 3-6, 7-10, 11-14 et 15-18. Il présente à des sujets 20 noms de chaque catégorie. Les sujets doivent donner une évaluation selon le label « caractère familier » à une

échelle de 1-5, avec 1 étant « peu familiers » et 5 étant « familiers ». Voici les données :

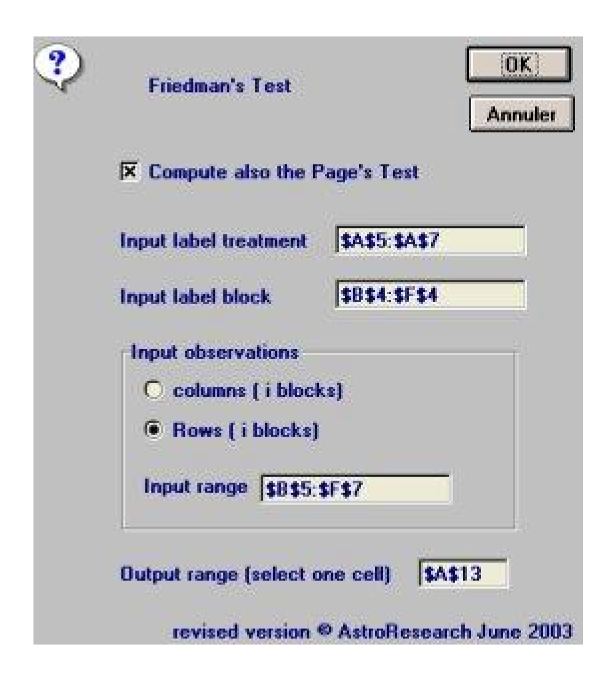
	3--6	**7--10**	**11--14**	**15--18**
1	5	5	4	5
2	5	5	4	5
3	5	5	5	4
4	5	5	5	5
5	5	5	5	4
6	5	5	5	3
7	5	5	5	2
8	4	5	5	2
9	5	5	5	4
10	5	4	5	3

Les résultats sont intéressants : ils montrent l'importance qu'il y a à prendre en compte les ex-aequo. Le test est en effet non significatif ; mais comme il y a 29 ex-aequo, en fait on trouve que p < 0.05.

Friedman's Test				1	2	3	4
N	40	3–6		3	3	3	2.5
N (=t)	10	7–10		3	3	3	2.5
k (=b)	4	11–14		1	1	3	2.5
SRj2	2624.5	15–18		3	3	1	2.5
Fr (3, 27)	6.38						
p	0.00206						
critical value 5%					3–6	7–10	11–14
ties	29		7–10		– 0 | 27.92		
corr. for ties	0.6		11–14		– 2.5 | 27.92	– 2.5 | 27.92	
T'	12.45		15–18		– 13.5 | 27.92	– 13.5 | 27.92	– 11 | 27.92
p'	0.00599						

TABLEAU XVI
(*statistique de Friedman - problème des ex aequo*)

L'interface de cette macro mérite quelques précisions. Le test de Friedman est assimilable à une ANOVA non paramétrique à deux facteurs : un facteur dit « *traitement* » et un facteur « *bloc* ». Dans le cas présent, le facteur bloc consiste en les 4 classes d'âge [les moyennes sont respectivement de : 29 - 29 - 26.5 - 15.5]. Bien que NS en formulation bilatérale, le test est significatif en formulation unilatérale. Le facteur traitement est ici de **N** = 10. Voici la boîte de dialogue permettant de gérer les données :



Ici, il convient donc de saisir comme bloc les intervalles d'âge [4] et comme traitement, les 10 mesures. Voici les formules permettant la computation de ce test. Il est incompréhensible que des revues de CD classiques distinguées - comme la revue ***Répertoire***, que j'évoque dans la section sur Sergiu Celibidache - n'aient pas employées cette statistique dans leur « *blindfold test* » lors des écoutes comparatives de versions différentes d'une oeuvre. Ce test permet de décider si les différences observées entre **t** échantillons répartis en **p** blocs sont dues au hasard, ou au fait que les traitements sont différents. Ce test s'applique donc surtout pour les expériences en blocs complètement randomisés. On commence par remplacer les observations par leur rang. L'opération s'effectue indépendamment pour chaque bloc. Ce test distribue les données en un tableau à double entrée ayant **N** rangées et ***k*** colonnes. Les rangées représentent les différents sujets [traitements = treatment] et les colonnes les différentes conditions [blocs = block]. Les données sont rangées. La détermination des rangs se fait pour chaque rangée séparément [cette méthode de classement par rangée séparée explique que nous ayons regroupés les statistiques de Friedman, de Page et le coefficient W de Kendall]. Donc pour ***k*** conditions, les rangs de chaque rangée se répartissent entre 1 et **k**. Le test détermine si les différentes colonnes de rangs proviennent de la même population. Le traitement des ex-aequo à l'intérieur d'un bloc reste le même que pour le test de Kruskal-Wallis. Etapes :

- détermination de **S**

$$S = \sum_{j=1}^{k} \left(R_{Tj} - n(k+1)/2\right)^2$$ (28)

Dans l'exemple du tableau XV, on trouve S = 1686, avec **R_{Tj}** = {27 ; 20 ; 19 ; 14}. **k** = 8 et **n** = 4.

- la statistique de Friedman est alors donnée par une formule qui suit une loi du chi carré à **k-1** ddl [df].

$$\chi_r^2 = \frac{12\sum_{j=1}^{k} \left(R_{Tj} - n(k+1)/2\right)^2}{nk(k+1)}$$ (29)

Notons que pour des échantillons de taille réduite, une formule différente est proposée, basée sur la statistique de Fisher [qui doit se lire à k-1 ; (**k -1**) x (**N -1**) ddl] :

$$Fr = \frac{(N-1)\cdot\left[\frac{\sum_{i=1}^{k} T_i^2}{N} - \frac{N\cdot k\cdot(k+1)^2}{4}\right]}{\sum(r_{ij}^2) - \frac{\sum_{i=1}^{k} T_i^2}{N}}$$ (30)

où comme précédemment, ST_i^2 est la somme des rangs pour chaque colonne [attention à ne pas confondre les « blocs » et les « traitements » !], **N** est le nombre de traitements et **k** le nombre de blocs. Sr_{ij}^2 est la somme des carrés de chaque rang des colonnes. On en voit un exemple au tableau XVI [à noter que la correction pour les ex aequo a été déterminée par le test du χ^2 à **k -1** ddl].

Mais cette formule ne tient pas compte des ex-aequo. On peut trouver cette formule à l'adresse suivante : http://ff123.net/.

$$T' = \frac{12\sum_{j=1}^{t}(x_{.j} - G/t)^2}{bt(t+1) - [1/(t-1)]\sum_{i=1}^{b}\left[\left(\sum_{j=1}^{g_i} t_{i,j}^3\right) - t\right]} \quad (31)$$

L'auteur fait référence à : ***Sensory Evaluation Techniques***, *3rd Ed.*, Meilgaard, Civille, and Carr). On remarque que les symboles diffèrent. Voici les correspondances : $\mathbf{R_{Tj}}$: $\mathbf{x_{.j}}$ **- n(k+1)/2** : **G - b** : **t - t** : **k**. $\mathbf{g_i}$ représente le nombre d'ex aequo du bloc **i** et $\mathbf{t_{i,j}}$ le nombre de valeurs des ex aequo **j** appartenant au groupe **i**. Si le test de Friedman conclut à un résultat globalement significatif, on peut tester la signification de comparaisons choisies, en étudiant les différences entre les sommes ou moyennes de rangs des traitements. On juge significatives des différences de sommes de rangs supérieures à la plus petite différence significative données par :

$$LSD_{rank} = t_{\alpha/2,\infty}\sqrt{bt(t+1)/6} \quad (32)$$

où $\mathbf{LSD_{rank} = |R_i - R_k|}$. Notez que Tomassone, dans son ***Biométrie***, rapporte le calcul de la différence des rangs à la loi Normale, alors que l'auteur la rapporte à une loi de Student Fisher. On a alors a' = a**/I(I-1)**, avec **I** : nombre de colonnes.

9bis) - le test de Page

Il est congénère du test de Jonkheere - Terpstra, à ceci près qu'on l'emploie dans le cas d'échantillons dépendants - au lieu que la statistique de Jonkheere est à employer pour des échantillons indépendants où l'on souhaite observer s'il existe une « structure ». Exemple : on ajoute un additif à un produit alimentaire ; ou bien encore, on ajoute un excipient à un médicament en sorte d'améliorer le goût d'une forme orale à saveur particulièrement désagréable [surtout chez l'enfant]. On souhaite apprécier l'effet de cette dose sur le jugement. C'est dans ces cas que la statistique de Page est conseillée. Là encore, nous allons donner quelques extraits d'une page rédigée en italien, tirée de : http://www.dsa.unipr.it/soliani/soliani.html.

« *Il test, detto anche per le alternative ordinate fondate sulla somma dei ranghi di Friedman (distribution-free test for ordered alternatives*

based on Friedman rank sums) è stato proposto da E. B. Page nel 1963 (con l'articolo ***Ordered hypotheses for multiple treatments: a significance test for linear ranks*** *pubblicato sulla rivista Journal of the American Statistical Association Vol. 58, pp. 216-230). Esso serve per verificare l'ipotesi che le mediane dei gruppi a confronto siano ordinate secondo una sequenza specificata a priori. Il test è solo unilaterale e quindi occorre definire a priori l'ordine naturale dei valori delle mediane.* »

La statistique de Page ressortit de l'hypothèse nulle $\mathbf{H_0}$:

$$H_0: me_1 = me_2 = me_3 = \ldots = me_K \qquad (33)$$

et de l'hypothèse alternative $\mathbf{H_1}$:

$$H_1: me_1 \leq me_2 \leq me_3 \leq \ldots \leq me_K \qquad (34)$$

exactement comme le test de Jonckheere, dans le cas d'échantillons dépendants. Elle est définie par **L**, où :

$$L = \sum_{j=1}^{k} j R_j \qquad (35)$$

L représente la somme des **j** rangs **Rj**, de **j** = 1 à **k**, où **k** est le nombre de « blocs » exactement comme dans le test de Friedman. La computation de **L** est donc identique à celle du test de Friedman ; nous l'avons lié, du reste, sur notre macro, au test de Page. Une table spéciale doit être employée pour observer la valeur critique de L à partir de laquelle $p < 0.05$. Voyons ceci sur un premier exemple - commun d'ailleurs à Tomassone et à Lamberto Soliani.

	A	**B**	**C**	**D**	**E**
1	2	1	4	5	3
2	2	1	3	5	4
3	1	3	2	4	5

Dans une expérience, on souhaite vérifier que l'apport d'une substance x, à 5 doses régulièrement espacées, entraîne un jugement de plus en plus mauvaix sur la qualité du produit. L'hypothèse à tester est donc :
H_1 : A ^{3} **B** ^{3} **C** ^{3} **D** ^{3} **E**, où au moins une inégalité est stricte. Nous donnons les résultats des tests de Friedman

et de Page :

Friedman's Test
N 15
t 3
b 5
SRj2 471
T 8.8
p 0.0663
critical value 5%
ties no
corr. for ties 1
T' 8.8
p' 0.0663

Page's Test
Page's Test
N (= t) 3
k (= b) 5
L 158
m 135
s 8.66
SL 2.656
p 0.05159 | exact p < 0.01
critical value for L [5 | 1 | 0.1] % 150 | 155 | 160

La boîte de dialogue appelant le test est la même que celle du test de Friedman, remaniée. Dans le cas où le nombre de blocs k est inférieur à 10, on consulte une table spéciale ; dans le cas présent, est supérieure à la valeur critique au seuil de 1 % [**L = 158155**]. Le **Z** [SL, qui correspond à $\mathbf{Z}^{0.5}$] est approximatif dans le cas où k est inférieur à 10 et ne doit pas être employé [il est donné ici à titre documentaire]. Au-delà, la formule à employer est :

$$L^* = \frac{L - \left(\frac{nk(k+1)^2}{4}\right)}{\sqrt{\frac{n(k^3-k)^2}{144(k-1)}}} \qquad (36)$$

où L* est l'approcimation normale de L, dès que k > 10 ;

elle suit une loi de χ^2 à 1 ddl. n est le nombre d'observations [il doit y avoir le même nombre à chaque colonne] et k est le nombre de « blocs ». Voyons un autre exemple, pour un grand échantillon : je le tire de l'ouvrage de Lamberto Soliani ; laissons s'exprimer l'auteur :

« *ESEMPIO 1 (PER GRANDI CAMPIONI). Lungo il corso d'acqua che*

attraversa una città, sono state collocate 6 stazioni (A, B, C, D, E, F) di rilevazione dell'inquinamento. In ognuna delle 6 stazioni, per 15 giorni è stata fatta una misura del carico inquinante. I valori campionati, classificati per stazione e per giorno di rilevazione, sono riportati nella tabella a due entrate sottostante: » [capu18.pdf, p. 1486]

	A	**B**	**C**	**D**	**E**	**F**
1	20	18	24	22	29	38
2	32	37	34	31	39	38
3	18	23	19	25	23	26
4	9	7	14	11	12	11
5	29	37	32	59	40	45
6	38	25	27	47	45	45
7	8	15	7	12	15	13
8	18	13	22	26	23	22
9	32	36	37	35	48	40
10	23	25	26	25	32	56
11	6	8	12	9	10	10
12	24	18	20	27	25	27
13	13	18	14	14	19	26
14	18	26	19	19	29	32
15	14	12	25	56	54	75

étape 1 : Nous devons d'abord ranger, séparément, ces observations comme pour le test de Friedman :

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**total R_i**
A	2	2	1	2	1	3	2	2	1	1	1	3	1	1	2	25
B	1	4	3.5	1	3	1	5.5	1	3	2.5	2	1	4	4	1	**37.5**
C	4	3	2	6	2	2	1	3.5	4	4	6	2	2.5	2.5	3	**47.5**
D	3	1	5	3.5	6	6	3	6	2	2.5	3	5.5	2.5	2.5	5	**56.5**
E	5	6	3.5	5	4	4.5	5.5	5	6	5	4.5	4	5	5	4	**72**
F	6	5	6	3.5	5	4.5	4	3.5	5	6	4.5	5.5	6	6	6	**76.5**

étape 2 : Calculer la moyenne **μ**, l'écart type **σ** et la statistique **Z** :

moyenne: $$\mu_L = \frac{N \cdot k \cdot (k+1)^2}{4}$$ (37),ici**μ = 1102.5 [**on rappelle que N = 15 et k = 5]

écart type : $\sigma_L = \sqrt{\frac{N \cdot (k^3 - k)^2}{144(k-1)}}$ (38), ici s = 30.311

Et enfin, $Z_L = \frac{12 \cdot L - 3 \cdot N \cdot k \cdot (k+1)^2}{k \cdot (k^2 - 1)} \cdot \sqrt{\frac{k-1}{N}}$ (39),
avec $\mathbf{Z}^{0.5}$ = 6.103, soit p = 0.00675.

Résultat sous Excel :

Page's Test

N (= t) 15
k (= b) 6
L 1287.5
m 1102.5
s 30.311
Z 6.103
p 0.00675
critical value for L [5 | 1 | 0.1] % not available

Dernier point : la formulation du test est, par essence, unilatérale [Z = 1.645, avec p < 0.05].

9 ter - coefficient de concordance de Kendall W

Là encore, on range les observations d'une façon semblable à ce qui est réalisé pour la statistique de Friedman. Il s'agit d'une sorte de généralisation à k classes [les blocs dans Friedman] du coefficient t de Kendall. W exprime donc le degré de concordance entre k classes. Pour calculer ce test, on procède ainsi :

- étape 1 : on classe les observations en rang, séparément pour chaque observation [il s'agit des étiquettes « traitement » dans Friedman] ;

- étape 2 : on calcule la somme des rangs j $S\mathbf{R}_j$ de chaque colonne **n** x **k** [**k** = classe = bloc et **N** = nombre d'observations = traitements ; il va de soi que le nombre d'observations doit être égal pour chaque classe, sinon le test n'a aucun sens].

- étape 3 : on détermine **W** :

$$W = \frac{12\sum_{i=1}^{N} R_i^2 - 3k^2 N(N+1)^2}{k^2 N(N^2-1) - k\sum_{j=1}^{k} T_j} \quad (40)$$

Il convient de tenir compte des ex aequo, qui diminuent un peu la variance de W ; le terme soustractif du dénominateur de **W** peut être déterminé par :

$$T_j = \sum_{i=1}^{g_j} (t_i^3 - t_i) \quad (41)$$

oùt_i, comme dans le test de Friedman, est le nombre d'ex aequo du groupe **i** de chaque colonne **j** ; g_j correspond, précisément, au nombre total de ces ex aequo dans le groupe **j**. Le degré de signification de **W** peut être lu dans une table du χ^2, pour la valeur :

$$X^2 = k(N-1)W \quad (42)$$

pour**N** -1 ddl [df]. Voyons cela sur un exemple. Supposons que l'on demande à trois mélomanes d'une revue X d'écouter 6 versions différentes d'une symphonie de Beethoven [mettons la 3ème, l'***Héroïque***] et de les ranger séparément suivant l'organisation des plans sonores [qui ressortissent de l'oganisation spatiale des instruments, laquelle varie en général grandement selon le chef d'orchestre]. Les trois séries indépendantes de rangs données par les trois mélomanes A, B, C sont exposées dans le tableau suivant :

	a	**b**	**c**	**d**	**e**	**f**
A	1	6	3	2	5	4
B	1	5	6	4	2	3
C	6	3	2	5	4	1

Nous pratiquons d'abord le test de Friedman :

Friedman's Test

N	18
t	3
b	6

SRj2 687
T 2.429
p 0.78715
critical value 5%
ties no
corr. for ties 1
T' 2.429
p' 0.78715

avec le détail des rangs, base de la computation du coefficient **W** de Kendall :

	A	**B**	**C**	**total R_i**
a	1	1	6	8
b	6	5	3	14
c	3	6	2	11
d	2	4	5	11
e	5	2	4	11
f	4	3	1	8

Kendall's coefficient of concordance

k 3
N 6
df 5
s 25.5
W 0.162
F 0.162
T for ties 0
W* 0.162
p 0.787
critical value of W [5 | 1] % |

On voit que le test n'est pas significatif ; les mélomanes n'ont pu départager les différentes versions de l'***Héroïque*** ! Nous allons voir un autre exemple, avec des ex aequo :

tableau des données :

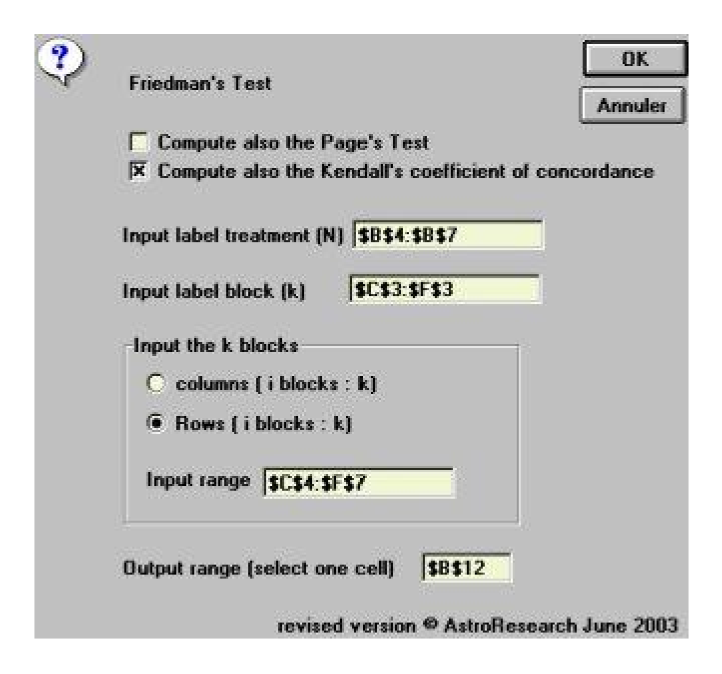
	A	**B**	**C**	**D**
I	1	3	3	3
II	1	4	2	3
III	2	3	1	4
IV	1.5	1.5	3.5	3.5

transformation en rangs et S**R**

	I	**II**	**III**	**IV**	**total Ri**
A	1	1	2	1.5	5.5
B	3	4	3	1.5	11.5
C	3	2	1	3.5	9.5
D	3	3	4	3.5	13.5

Notez que la transposition est faite automatiquement par le logiciel, à partir du moment où l'on a saisi la colonne

des étiquettes du tableau comme « traitements » [i.e. observations] et la ligne des étiquettes dudit tableau comme « blocs ». Voici du reste la boîte de dialogue qui correspond à la saisie de ce tableau.



Résultat :

Kendall's coefficient of concordance

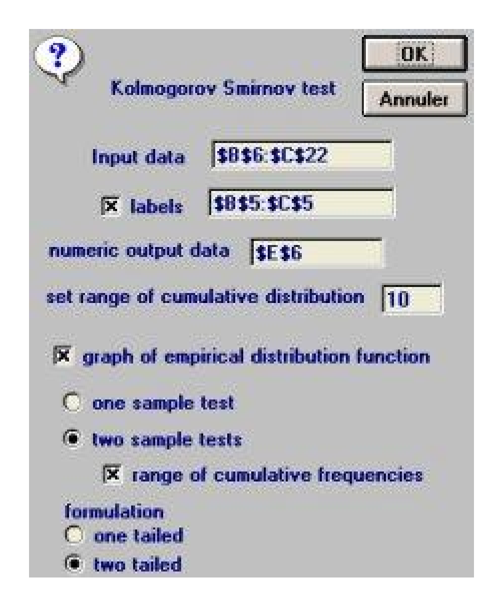
k 4
N 4
df 3
s 35
W 0.438
F 0.515
T for ties 3
W* 0.515
p 0.103
critical value of W [5 | 1] % |

10)- le test de Kolmogorov Smirnov è ne figure pas dans l'*Analysis Tool Pack* d'Excel

C'est une alternative des plus intéressantes au test du $\chi 2$ quand celui-ci n'est pas possible, au cas où les effectifs théoriques ne sont pas suffisants. Mais les propriétés de ce test sont beaucoup plus étendues que celles du χ^2. Ce test, en fait, est un véritable « *couteau suisse* » qui permet aussi bien de décider entre 4 auditeurs, quel est le meilleur violon entendu lors d'un « *blindfold test* », ou encore de décider si une courbe présente telle ou telle tendance [normale, . Aussi peut-on paraître un peu

étonné, en néophyte, des qualités protéiformes de ce te st.

Mise en oeuvre : elle s'établit en appelant la macro correspondante qui donne accès à cette boîte de dialogue :



- de haut en bas : saisie d'une ou de deux colonnes de données, selon que l'on traite un ou deux échantillons, cf. exemples infra. Saisie éventuelle des étiquettes des données [qui doivent figurer dans la première ligne] ; saisie de la cellule où l'analyse prendra forme. Puis : saisie du « facteur classe » pour l'établissement des classes de la fonction de répartition empirique, conduisant, sur option, à la construction d'un graphe. Chois de la modalité : un ou deux échantillons. Dans le cas où l'on dispose de deux échantillons de taille différente, il est obligatoire de classer les distributions cumulées [range of...]. Enfin, formulation uni ou bilatérale.

Voyons tout cela sur un premier exemple.

a)- <u>cas de deux échantillons</u>

TRT 0.38 1.26 0.34 0.7 1.75 50.57 1.55 0.08 0.42 0.5 3.2 0.15 0.49 0.95 0.24 1.37 0.17 6.98 0.1 0.94
placebo 1.19 2.37 2.16 14.82 1.73 41.04 0.23 1.32 2.91 39.41 0.11 27.44 4.51 0.51 4.5 0.18 14.68 4.66 1.3 2.06

Le cas de figure est le suivant : comparaison de deux modalités thérapeutiques, avec des tailles < 30. Cet exemple est extrait du site : http://www.physics.csbsju.edu/stats/KS-

test.n.plot_form.html

Laissons parler l'auteur :

« *The Kolmogorov-Smirnov test (**KS-test**) tries to determine if two datasets differ significantly. The KS-test has the advantage of making no assumption about the distribution of data. (Technically speaking it is non-parametric and distribution free.) Note however, that this generality comes at some cost: other tests (for example Student's t-test) may be more sensitive if the data meet the requirements of the test. [...]* »

L'intérêt de ce test est de mettre en évidence une différence significative, là où le test t de Student Fisher ne permet pas de conclure. Voici par exemple la conclusion du test t de l'***Analysis Tool Pack*** :

t-Test: Two-Sample Assuming Equal Variances

	Traitement	**Placebo**
Mean	3.607	8.3565
Variance	124.649148	164.336582
Observations	20	20
Pooled Variance	144.492865	
Hypothesized Mean Difference	0	
df	38	
t	-1.24946671	
P(T<=t) one-tail	0.10956794	
t Critical one-tail	1.68595307	
P(T<=t) two-tail	0.21913589	
t Critical two-tail	2.02439423	

Quelle est la procédure à suivre pour calculer la statistique de Kolmogorov - Smirnov ? Ici, nous avons deux échantillons. Nous devons :

- ranger les deux séries d'observations en fréquences cumulées [ce qui peut se faire de deux façons, selon que l'on distingue des classes ou pas ; en ce cas, nous calculons ce que l'on appelle la fonction de répartition empirique]. A propos de la fonction de répartition empirique, voici un extrait du site http://conte.montesquieu.u-bordeaux.fr/index.html où des précisions fort intéressantes sont mises en exergue :

« *La fonction de répartition est la fonction qui associe à tout nombre réel le nombre d • individus de la population étudiée qui ont une valeur inférieure au dit réel. Par inférieure, on entend en France strictement inférieur, alors que dans les pays anglo-saxons on entend inférieure ou égale. Dans ce qui suit, donc, si F est la fonction de répartition, pour tout x de **R**, l • image F(x) est l • effectif de la sous-population formée par les individus dont la valeur est strictement inférieure à x. [...] La représentation graphique (diagramme intégral) est alors un escalier, chaque marche étant ouverte à gauche et fermée à droite, la première étant de longueur infinie de* -oo *à* x_1 *et la dernière de longueur infinie de* x_n *à* +oo. »

- calculer la différence, en valeur absolue, de la quantité :

$$D_{n_1,n_2} = \sup_x |F_{n_1}(x) - G_{n_2}(x)|$$ (43)

Fn1 (x) symbolise le vecteur de fréquences cumulées de la variable x - ici le traitement - et Gn2 (y), le vecteur de la variable y - ici le placebo. On détermine la position de la différence maximale, en valeur absolue, des deux vecteurs.

- on relève, sur une table appropriée, la valeur critique de $D_{n1,n2}$ - ici dans une table pour deux échantillons.

Exemple, dans le cas présent :

F_{n1} (x)	**G_{n2} (x)**	**diff.**	**$D_{n1,n2}$**
0.00526754	0.00712021	0.81	0.00185267
0.02273357	0.02130078	1.11	0.00143279
0.02744663	0.03422485	1.82	0.00677822
0.03714999	0.12289834	14.12	0.08574836
0.06140837	0.13324957	0.02	0.07184119
0.76240643	0.37880692	9.53	0.38359952
0.78389243	0.38018309	1.32	0.40370934
0.78500139	0.38808113	1.24	0.39692025
0.7908234	0.40549273	2.49	0.38533067
0.79775437	0.64129719	38.91	0.15645717
0.84211256	0.64195536	3.09	0.20015719
0.84419185	0.80613893	27.29	0.03805292
0.8509842	0.83312392	4.02	0.01786028
0.86415304	0.83617543	0.44	0.0279776
0.8674799	0.86310058	4.26	0.00437932
0.88647075	0.86417759	1.19	0.02229317
0.88882728	0.9520134	14.51	0.06318612
0.98558359	0.97989589	2.32	0.0056877
0.98696978	0.98767427	1.2	0.00070448
1	1	1.12	0

On voit ici que **$D_{n1,n2}$** = 0.403. On remarque évidemment que ce tableau est pour le moins rébarbatif ; on peut pallier cette austérité en mettant en oeuvre un graphique approprié. Ce graphique n'est autre que celui qui illustre

la fonction de répartition empirique associée à un échantillon $x_1, \ldots x_n$. Cette fonction est définie par :

$$F_n : R \to [0,1]$$

$$x \mapsto F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \le x\}} = \begin{cases} 0 & \text{si } x < x_1^* \\ i/n & \text{si } x_i^* \le x < x_{i+1}^* \\ 1 & \text{si } x \ge x_n^* \end{cases} \quad (44)$$

On en voit le résultat dans les colonnes de gauche du tableau [F_{n1} (x) et G_{n2} (x)]. La fonction de répartition empirique est une fonction en escalier qui fait des sauts de hauteur **1/n** en
chaque point de l • échantillon. Elle est donc dite « ouverte à gauche » et « fermée à droite. »

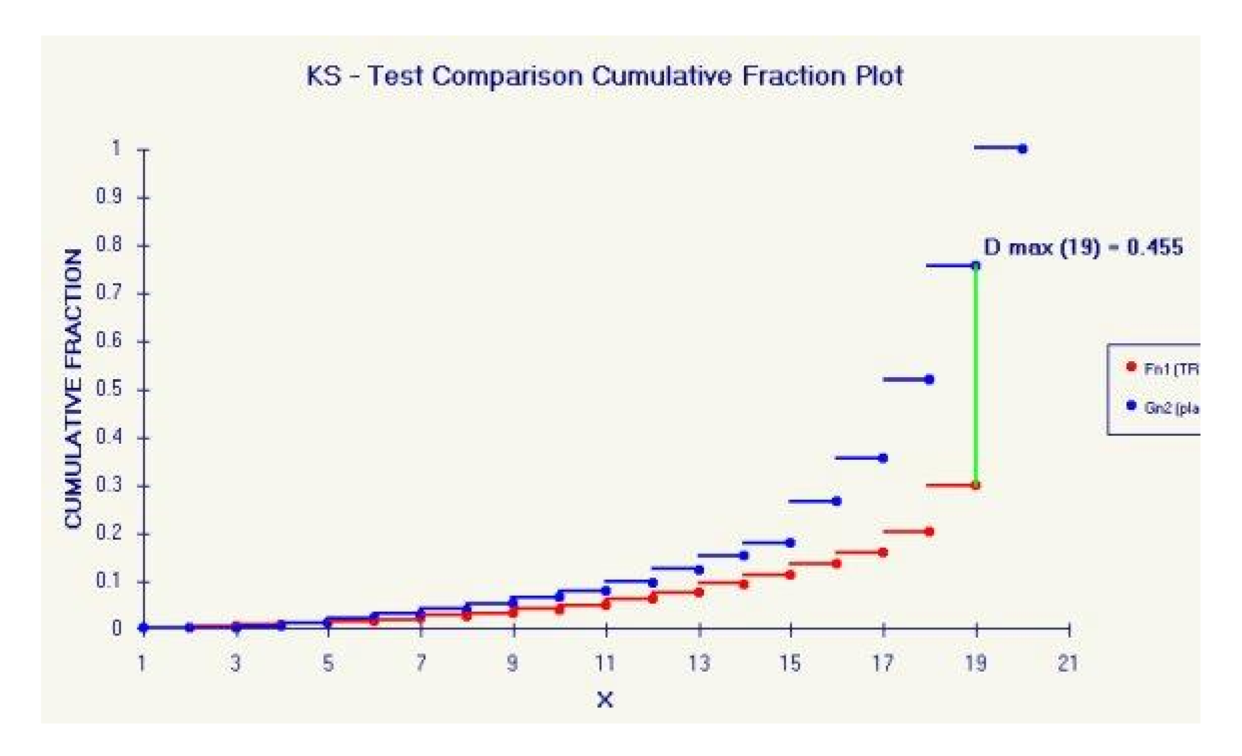


FIGURE II
(fonction de répartition empirique)

La série colorée en rouge est le traitement, celle en bleu est le placebo : de façon générale, la série de gauche est donc en **rouge** et la série en **bleu** est disposée à droite. Le tableau suivant montre l'analyse du test de Kolmogorov Smirnov pour deux échantillons :

Kolmogorov-Smirnov for two samples	
N	40
n1	20
n2	20
S1	72.14
S2	167.13
m1	3.607
m2	8.357
Dk	0.404
Kd	8
pD	p < 0.05
0.05	0.4
0.01	0.5
pKd	p < 0.05
0.05	8
0.01	10

TABLEAU XVII

Nous avons accès aux informations suivantes : taille N d l'échantillon ; tailles n1 et n2 des colonnes ; sommes et moyennes ; puis deux indices qui forment le résultat du test, exprimés par Dk et Kd. Ici, la valeur **Dk = 0.404** est significative, avec p < 0.05 puisqu'elle est supérieure à la valeur critique 0.4, de justesse. C'est ce qui est exprimé sur le graphique : la barre verte verticale indique le point où l'on a trouvé la valeur **SUP**de$\mathbf{D}_{n1,n2}$ [ici, le point 7] avec la valeur correspondante. Les valeurs critiques peuvent être trouvées dans les tables appropriées et ce sont les indications qui apparaissent plus bas, avec les valeurs critiques au seuil de 5% et de 1%. Ces tables peuvent être trouvées, pour partie, sur le site suivant : http://cons-dev.univ-lyon1.fr/Enseignement/Stat/

Il est nécessaire de disposer des tables suivantes :

- valeurs critiques pour un échantillon [Siegel S (1956), Non-parametric Statistics for the Behavioral Sciences McGraw • Hill] ;
- idem pour deux échantillons, avec n1 = n2 pour n < 40 [Sprent, 1992] et n > 40 [Siegel, 1956] ;
- idem pour deux échantillons avec n1 <> n2 ;
- grands échantillons, bilatéral [Siegel, 1956].

Ces tables sont intégrées dans la feuille macro ; elles peuvent être consultées et copiées.

Mais il y a une autre manière de regrouper les distributions cumulées, en classes. On pourra trouver le détail de la façon de classer des distributions sur le site suivant dont nous tirons, là encore, cet extrait http://conte.montesquieu.u-bordeaux.fr/index.html :

« *Rappel : par définition,* Ni< *(ou)* Ni<= *est le nombre total d • individus ayant une valeur prise dans l'ensemble des i- 1 (ou i) premières modalités, la* $i^{ème}$ *modalité étant notée* $[x_{i-1}, x_i[$.

Dans un cumul à la française (ou à l'anglo-saxonne), on inscrit donc à la ligne i, en face de [x_{i-1}, x_i[, le nombre Ni< *(ou* Ni<=*), c'est-à-dire le nombre d'individus dont la valeur est inférieure à* xi-1 *(ou à* xi*).*
Insistons sur la manière de lire le tableau pour en extraire l'information disponible. Puisque à la ligne i, Ni< *est le nombre d'individus qui ont une valeur prise dans les i -1 premiers intervalles, la plus grande valeur possible d'un individu pris dans ces* Ni< *individus est donc au plus égale à la plus grande valeur du (i , 1)ème* intervalle, donc *inférieure à sa borne droite,* x_{i-1}. *Ainsi, la plus grande valeur possible des* Ni< *premiers individus (ligne i d'un cumul à la française) est donc inférieure à* x_{i-1}, *qui est la borne droite de la (i - 1)ème* modalité mais aussi la borne gauche de la *ième.*

De même Ni<= *est le nombre de ceux dont la valeur est au plus égale à la plus grande valeur du ième intervalle de valeurs, donc inférieure à sa borne droite,* x_i, *qui est aussi la borne gauche du i +1ème.*
Au total, les effectifs Ni< *sont décalés comme précédemment, vers le bas ou vers la droite.* »

Appliquons cela à notre exemple :

rank	**Fn1 (TRT)**	**Gn2 (placebo)**	**D n1,n2** <u>0.45</u>
[0 - 1 [	0.65	0.2	0.25
[2 - 3 [	0.85	0.6	0.15
[4 - 5 [	0.9	0.75	0.2
[6 - 7 [	0.95	0.75	0.2
[8 - 9 [	0.95	0.75	0.2
[10 - 11 [	0.95	0.75	0.2
[12 - 13 [	0.95	0.75	0.1
[14 - 15 [	0.95	0.85	0.1
[16 - 17 [	0.95	0.85	0.1
[18 - 19 [	0.95	0.85	0.1
[20 - 21 [	0.95	0.85	0.1
[22 - 23 [	0.95	0.85	0.1
[24 - 25 [	0.95	0.85	0.1
[26 - 27 [	0.95	0.85	0.05
[28 - 29 [	0.95	0.9	0.05
[30 - 31 [	0.95	0.9	0.05
[32 - 33 [	0.95	0.9	0.05
[34 - 35 [	0.95	0.9	0.05
[36 - 37 [	0.95	0.9	0.05
[38 - 39 [	0.95	0.9	0
[40 - 41 [	0.95	0.95	0.05
[42 - 43 [	0.95	1	0.05
[44 - 45 [	0.95	1	0.05
[46 - 47 [	0.95	1	0.05
[48 - 49 [	0.95	1	0
[50 - 51 [	1	1	

complété par le graphique suivant ; voici, à cette étape, les commentaires de l'auteur [http://www.physics.csbsju.edu/stats/KS-test.n.plot_form.html] :

« *You can see that the control and treatment datasets span much the same range of values (from about .1 to about 50). But for most any x value, the fraction of the treatment group that is strictly less than x is clearly less than the fraction of the control group that is less than x. That is, by-and-in-large the treatment values are larger than the control values for the same cumulative fraction. For example, the median (cumulative fraction =.5) for the control is clearly less than one whereas the median for the treatment is more than 1. The KS-test uses the maximum vertical deviation between the two curves as the statistic D. In this case the maximum deviation occurs near x=1 and has D=.45. (The fraction of the treatment group that is less then one is 0.2 (4 out of the 20 values); the fraction of the treatment group that is less than one is 0.65 (13 out of the 20 values). Thus the maximum difference in cumulative fraction is D=.45.)* »

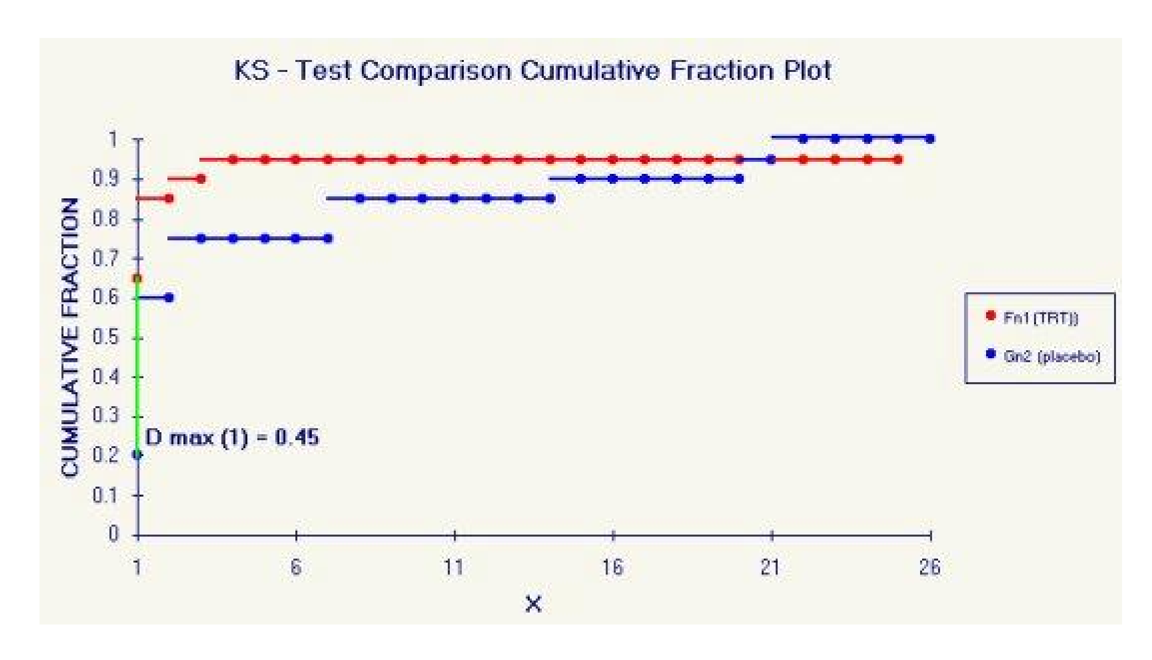


FIGURE III
(fonction de répartition empirique)

b)- cas d'un échantillon

exemple 1 : nous le tirons de l'ouvrage de Tomassone [cf. ref.]. Il s'agit d'un test de dégustation de vin dans des verres plus ou moins foncés. Voici les données de base :

couleur du verre de moins foncée à plus foncée	nombre de personnes désignant le vin comme le meilleur
1	0
2	1
3	0
4	5
5	4

et voici les résultats :

Fn (vin)	F (vin)	d
0.2	0	0.2
0.4	0.1	0.3
0.6	0.1	0.5
0.8	0.6	0.2
1	1	
Kolmogorov-Smirnov for one sample		
Dk	0.5	
n rank	5	
N	10	
m	2	
p	p < 0.01	
D critical 0.05	0.41	
D critical 0.01	0.49	

TABLEAU XVIII

Et les auteurs d'écrire :

« *nous pouvons conclure que la couleur du verre a une influence sur la sensation gustative. Ce qui fait mentir le vieux dicton : " qu'importe le flacon pourvu qu'on ait l'ivresse. " ...* »

En effet, on trouve **Dk = 0.5**, valeur supérieure au seuil critique à 1%, 0.49. La colonne de gauche Fn (vin) représente la fréquence théorique et F (vin) la fréquence observée. Dans la colonne de droite, d, la valeur maximale est 0.5 [point 3], ce qu'illustre la figure suivante :

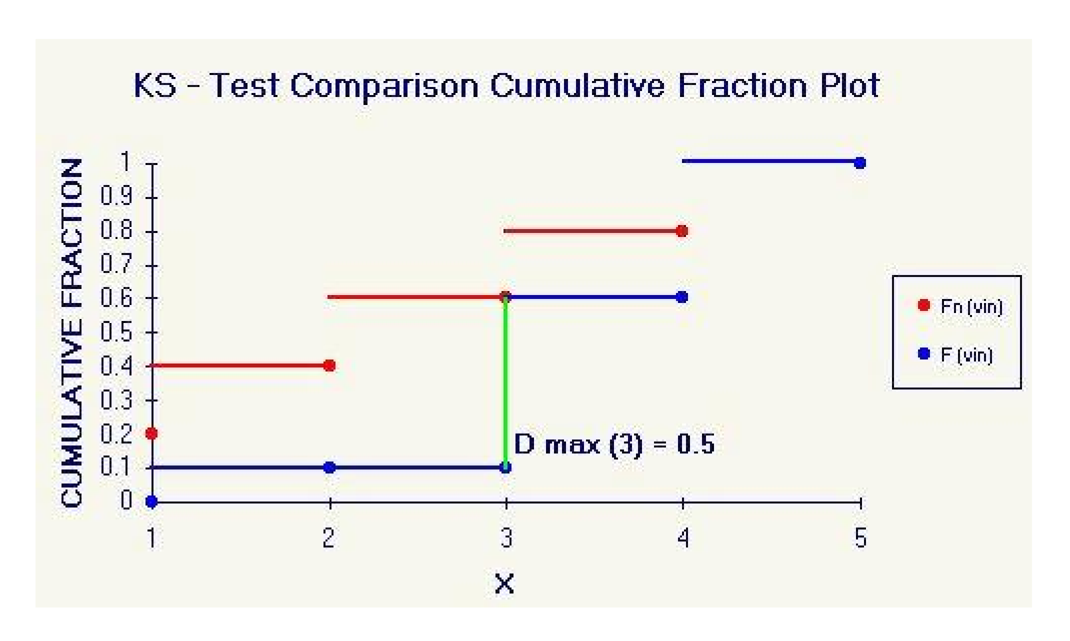


FIGURE IV
(fonction de répartition empirique)

c)- un dernier exemple

Il est tiré du site : http://www.physics.csbsju.edu/stats/KS-test.n.plot_form.html.

« *Two near-by apple trees are in bloom in an otherwise empty field. One is a Whitney Crab the other is a Redwell. Do bees prefer one tree to the other? We collect data by using a stop watch to time how long a bee stays near a particular tree. We begin to time when the bee touches the tree; we stop timing when the bee is more than a meter from the tree. (As a result all our times are at least 1 second long: it takes a touch-and-go bee that long to get one meter from the tree.) We wanted to time exactly the same number of bees for each tree, but it started to rain. Unequal dataset size is not a problem for the KS-test.* »

Voici comment se présentent les données brutes :

redwell = {23.4, 30.9, 18.8, 23.0, 21.4, 1, 24.6, 23.8, 24.1, 18.7, 16.3, 20.3, 14.9, 35.4, 21.6, 21.2, 21.0, 15.0, 15.6, 24.0, 34.6, 40.9, 30.7, 24.5, 16.6, 1, 21.7, 1, 23.6, 1, 25.7, 19.3, 46.9, 23.3, 21.8, 33.3, 24.9, 24.4, 1, 19.8, 17.2, 21.5, 25.5, 23.3, 18.6, 22.0, 29.8, 33.3, 1, 21.3, 18.6, 26.8, 19.4, 21.1, 21.2, 20.5, 19.8, 26.3, 39.3, 21.4, 22.6, 1, 35.3, 7.0, 19.3, 21.3, 10.1, 20.2, 1, 36.2, 16.7, 21.1, 39.1, 19.9, 32.1, 23.1, 21.8, 30.4, 19.62, 15.5}

whitney = {16.5, 1, 22.6, 25.3, 23.7, 1, 23.3, 23.9, 16.2, 23.0, 21.6, 10.8, 12.2, 23.6, 10.1, 24.4, 16.4, 11.7, 17.7, 34.3, 24.3, 18.7, 27.5, 25.8, 22.5, 14.2, 21.7, 1, 31.2, 13.8, 29.7, 23.1, 26.1, 25.1, 23.4, 21.7, 24.4, 13.2, 22.1, 26.7, 22.7, 1, 18.2, 28.7, 29.1, 27.4, 22.3, 13.2, 22.5, 25.0, 1, 6.6, 23.7, 23.5, 17.3, 24.6, 27.8, 29.7, 25.3, 19.9, 18.2, 26.2, 20.4, 23.3, 26.7, 26.0, 1, 25.1, 33.1, 35.0, 25.3, 23.6, 23.2, 20.2, 24.7, 22.6, 39.1, 26.5, 22.7}

En quelques étapes, nous allons traiter ces données :

1. transformation des données brutes en données tabulées : cf. infra la macro SPACE.
2. construction de la fonction de réparition empirique :

a. classement des données :

rank	Fn1 (redwell))	Gn2 (whitney)	D n1,n2
[1 - 2 [			
[3 - 4 [	0.09459459	0.07594937	0.01864522
[5 - 6 [	0.09459459	0.07594937	0.01864522
[7 - 8 [	0.09459459	0.07594937	0.01864522
[9 - 10 [	0.10810811	0.08860759	0.01950052
[11 - 12 [	0.10810811	0.08860759	0.01950052
[13 - 14 [	0.12162162	0.12658228	0.00496066
[15 - 16 [	0.12162162	0.17721519	0.05559357
[17 - 18 [	0.17567568	0.18987342	0.01419774
[19 - 20 [	0.22972973	0.25316456	0.02343483
[21 - 22 [	0.36486486	0.30379747	0.06106739
[23 - 24 [	0.58108108	0.36708861	0.21399247
[25 - 26 [	0.7027027	0.62025316	0.08244954
[27 - 28 [	0.81081081	0.78481013	0.02600068
[29 - 30 [	0.83783784	0.89873418	0.06089634
[31 - 32 [	0.83783784	0.94936709	0.11152925
[33 - 34 [	0.87837838	0.96202532	0.08364694
[35 - 36 [	0.90540541	0.97468354	0.06927813
[37 - 38 [	0.94594595	1	0.05405405
[39 - 40 [	0.95945946	1	0.04054054
[41 - 42 [	0.98648649	1	0.01351351
	1	1	0

b)- test

Kolmogorov-Smirnov for two samples		
N		153
n1		74
n2		79
S1		1586.22
S2		1667.9
m1		21.44
m2		21.11
Dk		0.214
Kd		16
pD		
	0.05	0.11
	0.01	0.132
pKd		p < 0.01
	0.05	
	0.01	

TABLEAU XIX

En formulation unilatérale, on observe que la valeur Dk est supérieure à la valeur critique 0.132.

c)- <u>graphique associé</u>

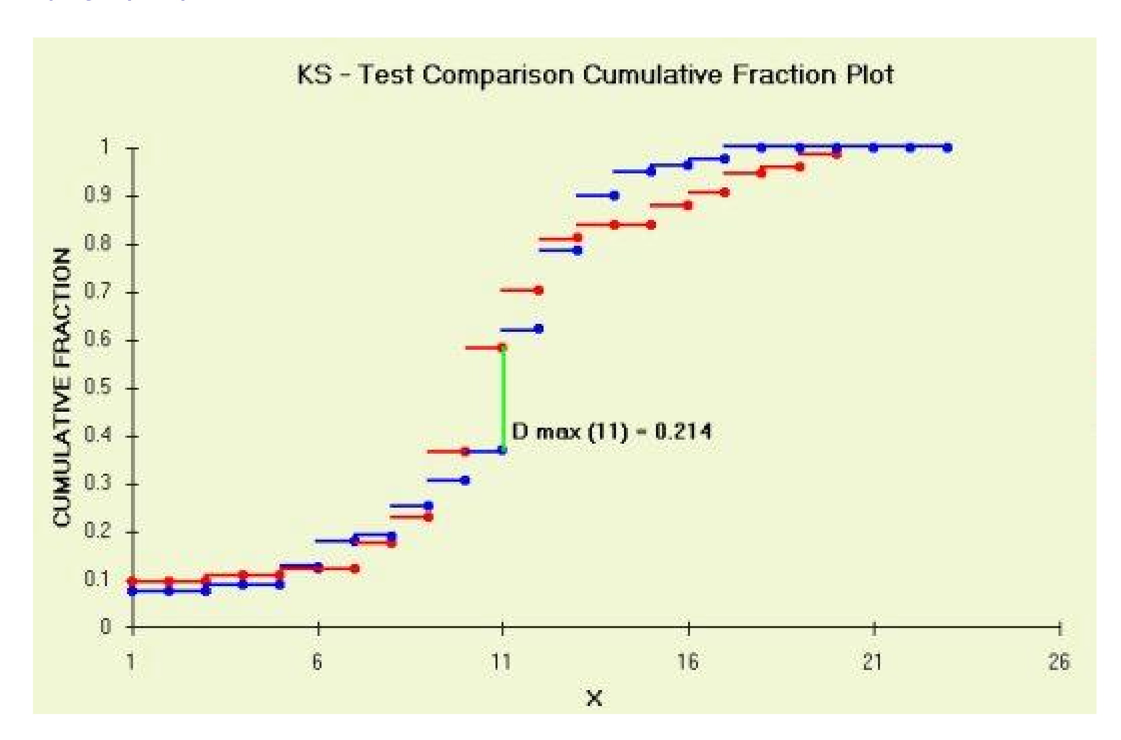


FIGURE V

(fonction de répartition empirique)

L'auteur conclue en ces termes :

« *This example is based on data distributed according to the Cauchy distribution: a particularly abnormal case. The plots do not look particularly abnormal, however the large number of outliers is a tip off of a non-normal distribution. The web page is silent on if this data is normal or lognormal; that means it finds no evidence for either possibility. This relatively large sample size can not save the t-test: it cannot see the difference, whereas the KS-test can. The t-test is not robust enough to handle this highly non-normal data with N=80.* »

Effectivement, le test t ne montre pas de différence significative :

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	Variable 2
Mean	21.4354054	21.1126582
Variance	91.3832279	65.8441967
Observations	74	79

Pooled Variance 78.1908807
Hypothesized Mean Difference 0
df 151
t 0.22561506
P(T<=t) one-tail 0.41090295
t Critical one-tail 1.65500751
P(T<=t) two-tail 0.8218059
t Critical two-tail 1.97579993

11. Tau de Kendall : la corrélation de rang

Ce test est congénère du coefficient de Spearman. Aussi a-t-il été intégré dans le logiciel avec le calcul de ce coefficient. Comme tous les autres tests relatifs à la corrélation, on doit dire avant tout que s'il s'agit d'un test qui met en évidence un certain degré de liaison entre deux variables, il ne peut en aucun cas assurer qu'il existe un lien de causalité entre ces deux variables. Le Tau - t - de Kendall est employé dans les cas ou les variables sont ordinales et discontinues ou encore dans les cas où les distributions des deux variables sont dites « non-normales ». En somme, Le tau de Kendall est un autre indicateur du caractère éventuellement significatif de la relation qui existe entre deux classements. Mais il possède un caractère singulier : il mesure la corrélation d'observations appariées, ce qui le rend, d'une certaine façon, semblable au test T de Wilcoxon.

La marche à suivre pour déterminer le t de Kendall est sensiblement la même que lorsqu'on classe les observations en rang, séparément, pour chaque colonne. On range ainsi les colonnes X et Y. Dans un second temps, on arrange le vecteur X par valeurs croissantes [sur Excel, il faut employer la commande TRIER et indiquer une seule clef de tri, en sorte que les valeurs du vecteur Y soient réarrrangéses en fonction de celles du vecteur X]. Puis pour tout couple d'items dont l'ordre dans le premier classement est concordant avec celui du second on attribue un score de + 1; dans le cas de discordance on attribue un score de - 1. **S** est la somme de ces scores sur l'ensemble des **n(n - 1)/2** comparaisons possibles. Le coefficient de corrélation de rangs de Kendall est défini par :

$$\tau = \frac{S}{C_n^2} = \frac{S}{\frac{1}{2}n(n-1)}$$ (45)

Voyons cela sur un premier exemple, tiré du site :
http://www.cohort.com/index.html.

« *Data for the sample run is from Sokal and Rohlf (Box 15.6, 1981; or Box 15.7, 1995): "Computation of rank correlation cœfficient between the total length (Y_1) of 15 aphid stem mothers and the mean thorax length (Y_2) of their parthenogenetic offspring.*" »

Y1	Y2
8.7	5.95
8.5	5.65
9.4	6
10	5.7
6.3	4.7
7.8	5.53
11.9	6.4
6.5	4.18
6.6	6.15
10.6	5.93
10.2	5.7
7.2	5.68
8.6	6.13
11.1	6.3
11.6	6.03

Voici le résultat obtenu sur Excel, au moyen de la macro spécifique au coefficient de rang de Spearman auquel est ajouté - par une case à cocher - le Tau de Kendall

Spearman rank order correlation

N 30
Np 15
Mx 9
My 5.74
Sd 0

Sd2 196.5
df 13
r' 0.649
t 3.076
p 0.00885
Z 0.00758
ties 2
r' corr. 0.649
t corr. 3.074
p corr. 0.00888

Kendall's Tau

n 15
S 52
tau 0.4952
Tk for ties X : 0 - Y : 1
tau corr. for ties 0.4976
p 0.0101
p corr. 0.0097
critical value tau (n<13) not available

Un graphique - le même que celui du coefficient de Spearman - est en outre disponible :

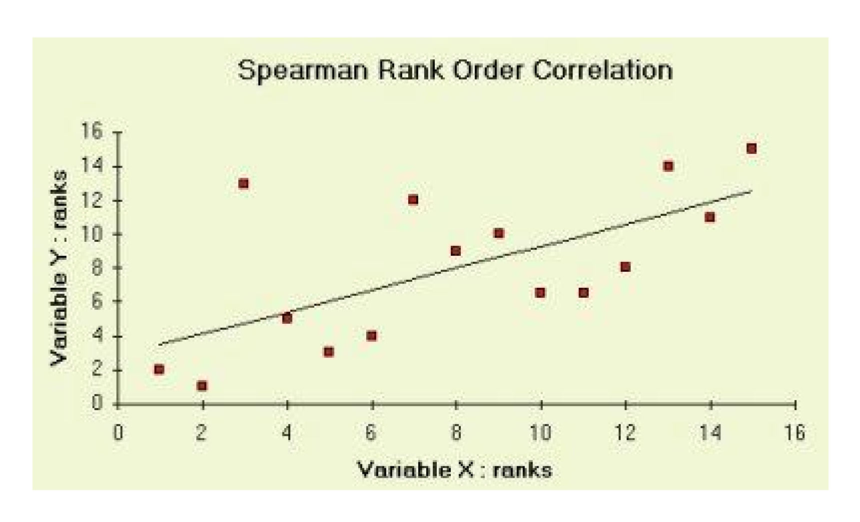


FIGURE VI

Les auteurs du site expriment ce résultat sous la forme suivante :

Y1 column: 1) Y1

Y2 column	n	Kendall tau	P	Spearman r	P
2) Y2	15	0.49761335153	(n<=40)	0.64910714286	.0088 **

P is the probability that the variates are not correlated. The low P value (<=0.05) for this data set indicates that the two variates probably are correlated.

Il y a 2 ex aequo ; la formule à employer devient :

$$\tau = \frac{S}{\sqrt{\frac{1}{2}n(n-1) - T_1}\sqrt{\frac{1}{2}n(n-1) - T_2}} \quad (46)$$

où T_1 et T_2 sont calculés ainsi :

$$T_k = \frac{1}{2}\sum_i t_i(t_i - 1) \quad (47)$$

où t_i est le nombre d'ex aequo au rang i pour la variable $\mathbf{X_k}$ (k = 1, 2). Dans le cas présent, $\mathbf{T_k}$ = 1 pour la variable **Y** [X_2]. La correction, comme on voit, est minime sur cet exemple. Le résultat est significatif puisque p = 0.01. Ici, **n** = 15 et **S** = 52 ; on peut donc employer l'approximation suivante :

$$z = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \quad (48)$$

qui suit une loi normale. Si **n** < 13 et **S** < 50, il faut disposer d'une table spéciale. Elle est intégrée au logiciel et consultable comme les autres. A noter que je n'ai pu réussir à la trouver sur le Net, malgré des recherches approfondies. J'ai donc pris cette table dans le Tomassone, référencé infra. Nous allons présenter un autre exemple, tiré du site :
http://cons-dev.univ-lyon1.fr/Enseignement/Stat/stat7/st7.html [notez que cette page n'est plus disponible qu'en cache, sur Google].

« *La relation entre l'autoritarisme des étudiants et leur conformisme social est recherché. L'autoritarisme des sujets et leur conformisme social sont appréciés par le passage de tests.* »

étudiant	conformisme	autoritarisme
A	82	42
B	98	46
C	87	39
D	40	37
E	116	65
F	113	88
G	111	86
H	83	56
I	85	62
J	126	92
K	106	54
L	117	81

Spearman rank order correlation

N 24
Np 12
Mx 97
My 62.33
Sd 0
Sd2 52
df 10
r' 0.818
t 4.497
p 0.00115
Z 0.00333
ties no
r' corr.
t corr.
p corr.

Kendall's Tau

n 12
S 44
tau 0.6667
Tk for ties X : 0 - Y : 0

tau corr. for ties 0.6667
p 0.0026
p corr. 0.0026
critical value tau (n<13) 0.001

Notez que la valeur critique s'entend au seuil unilatéral ; alors que nous avons calculé la valeur de **p** en formulation bilatérale. On peut aussi déterminer un coefficient de rang partiel τ_{xy}.Quand une corrélation est observée entre deux variables, il y a toujours la possibilité que cette corrélation soit due à l'association entre chacune des deux variables et une troisième variable. Les effets de variation due à une troisième variable sur la relation entre deux autres variables **X** et **Y** sont éliminés par une corrélation partielle. D'une autre façon, la corrélation entre **X** et **Y** est calculée alors que la troisième variable est maintenue constante. Nous espérons donner bientôt une suite pratique à cet exposé sur le coefficient de rang partiel.

Il est certain que le Tau de Kendall possède des avantages pratiques et pédagogiques. Les statisticiens calculent assez souvent une quantité comme un coefficient de corrélation sans demander ce que la quantité signifie, au sens de : « est-elle signifiante ? ». Car une différence peut être significative sans que pour autant elle soit le moins du monde signifiante. Le coefficient de Kendall, d'autre part, a une interprétation intuitivement simple. Qui plus est, sa structure algébrique est beaucoup plus simple que celui du coefficient de Spearman. Il peut même être calculé à partir des observations réelles sans passer par le classement en rang, se rapprochant un peu, en cela, du test de Komogorov Smirnov. Un coefficient de corrélation est destiné à mesurer « la force d'une liaison ». Mais des coefficients de corrélation différents mesurent une force de liaison de différentes façons. Seul le coefficient Kendall a une interprétation simple [adapté de ***Why Kendall Tau***? de G. E. NOETHER, University of Connecticut].

12. Test de la médiane [Mood test]

Il s'agit du test du signe, adapté à deux ou à plusieurs échantillons

« *This is a crude version of the Kruskal-Wallis ANOVA in that it assesses the difference in samples in terms of a contingency table. The number of cases in each sample that fall above or below the common median is counted and the Chi-square value for the resulting*

2 x k samples contingency table is calculated. Under the null hypothesis (all samples come from populations with identical medians), approximately 50% of all cases in each sample are expected to fall above (or below) the common median. The Median test is particularly useful when the scale contains artificial limits, and many cases fall at either extreme of the scale (off the scale). In this case, the Median test is the most appropriate method for comparing samples... »

Exemples :

soit deux groupes de livres de statistique et de livres divers ; on cherche à savoir s'ils sont paginés de la même façon. Voici les données

divers 29 39 60 78 82 112 125 170 192 224 263 275 276 286 756

stat 126 142 156 228 245 246 370 419 433 454 478 503 369

La médiane vaut **236.5**.

	divers	***stat***	
x < 236.5	10	4	14
x >= 236.5	6	8	14
	16	12	28

	divers	***stat***	**nij**
x < 236.5	0.5	0.67	1.17
x >= 236.5	0.5	0.67	1.17

nji 1 1.34 2.34

Contingency table

ddl 1
chi2 2.33
p 0.123

Le test exact de Fisher permet de trouver une valeur semblable

Possible matrices and their Ps

p for 0 : [14,**0**,2,12]

2.9913E-06

p for 1 : [13,**1**,3,11]

0.00016751

p for 2 : [12,**2**,4,10]

0.00299427

p for 3 : [11,**3**,5,9]

0.02395417

p for 4 : [10,**4**,6,8]

0.09881097

p for 5 : [9,**5**,7,7]

0.22585364

p for 6 : [8,**6**,8,6]

0.2964329

p for 7 : [7,**7**,9,5]

0.22585364

p for 8 : [6,**8**,10,4]

0.09881097

p for 9 : [5,**9**,11,3]

0.02395417

p for 10 : [4,**10**,12,2]

0.00299427

p for 11 : [3,**11**,13,1]

0.00016751

p for 12 : [2,**12**,14,0]

2.9913E-06

Fisher's exact test

table = [10,4,6,8]
one tailed p 0.12592991
two tailed p 0.15304886
other 0.97288
theor. value of chi2 < 3 6 for [4]

Le test de la médiane peut être étendu à plusieurs groupes [extension of the median test]. On peut trouver un exemple fort instructif au site suivant :

http://cons-dev.univ-lyon1.fr/Enseignement/Stat/St.html.Je laisserai le lecteur aller consulter la page.

Voici enfin les formules qui permettent d'opérer la computation du test, qui se résume en fait à un classement des différentes valeurs selon la médiane [inclue ou exclue selon le cas] suivi d'un test du chi2 classique ou d'un tableau de contingence. Ces formules sont extraites de: NPARTESTS.

$$Md = \begin{cases} \left(X_{[N/2]} + X_{[N/2+1]}\right)/2 & \text{if } N \text{ is even} \\ X_{[(N+1)/2]} & \text{otherwise} \end{cases} \qquad (49)$$

Voici un exemple ; il est tiré du site :http://cons-dev.univ-lyon1.fr/Enseignement/Stat/St.html.

«*Supposons qu'un chercheur veuille étudier l'influence du niveau d'instruction des mères sur le degré d'intérêt qu'elles présentent pour la scolarité de leurs enfants. Le niveau d'instruction de chaque mère est apprécié par le diplôme le plus important obtenu par chacune d'elle et leur degré d'intérêt pour la scolarité des enfants est mesuré par le nombre de visites volontaires que chacune d'elle rend à l'école. En tirant au hasard un nom sur dix d'une liste de 440 enfants inscrit à l'école, il obtient les noms de* ***44*** *mères, qui constituent son échantillon. Son hypothèse est que le nombre de visites varie en fonction du niveau d'éducation des mères.*
Comme les groupes de mères de divers niveau d'instruction sont indépendants les uns des autres et que plusieurs goupes sont formés, un test pour ***k*** *échantillons indépendants est envisagé. Comme le nombre d'années de scolarité des mères et que le nombre de visites constituent au mieux des mesures ordinales du niveau d'instruction et du degré d'intérêt, le test des médianes est considéré comme le mieux adapté à tester l'hypothèse*

concernant des différences des tendances centrales. »

Voici les données :

Primaire 4 3 0 7 1 2 0 3 5 1
Collège 2 4 1 6 3 0 2 5 1 2 1
Terminale 2 0 4 3 8 0 5 2 1 7 6 5 1
Université (1 ans) 9 4 2 3
Univ. (2) 2 4 5 2
Univ. (>2) 2 6

	Primaire	Collège	Terminale	Université (1	Univ. (2)	Univ. (>2)	
x < 2.5	5	7	6	1	2	1	22
x >= 2.5	5	4	7	3	2	1	22
	10	11	13	4	4	2	44

TABLEAU XX
(tableau de contingence du Mood Test)

Mais comme plusieurs effectifs théoriques sont inférieurs à 3, on ne peut traiter ce tableau par un $\chi 2$ d'homogénéité. On doit, soit regrouper des données, soit pratiquer un test de Kolmogorov Smirnov ; essayons cette option :

rank	Fn1 (x < 2.5)	Gn2 (x >= 2.5)	D n1,n2
[0 - 1.5 [	0.33333333	0.16666667	0.16666666
[1.5 - 2.1 [	0.5	0.33333333	0.16666667
[2.1 - 2.7 [	0.5	0.33333333	0.16666667
[2.7 - 3.4 [	0.5	0.5	0
[3.4 - 4 [	0.5	0.66666667	0.16666667
[4 - 4.6 [	0.5	0.83333333	0.33333333
[4.6 - 5.3 [	0.66666667	1	0.33333333
[5.3 - 5.9 [	0.66666667	1	0.33333333
[5.9 - 6.5 [	0.83333333	1	0.16666667
[6.5 - 7.2 [	1	1	0

Kolmogorov-Smirnov for two samples
N 12
n1 6
n2 6
S1 22
S2 22
m1 3.67
m2 3.67
Dk 0.333
Kd 1
pD NS
D critical 0.05 0.833
D critical 0.01 1
pKd NS

Kd critical 0.05 5
Kd critical 0.01 6

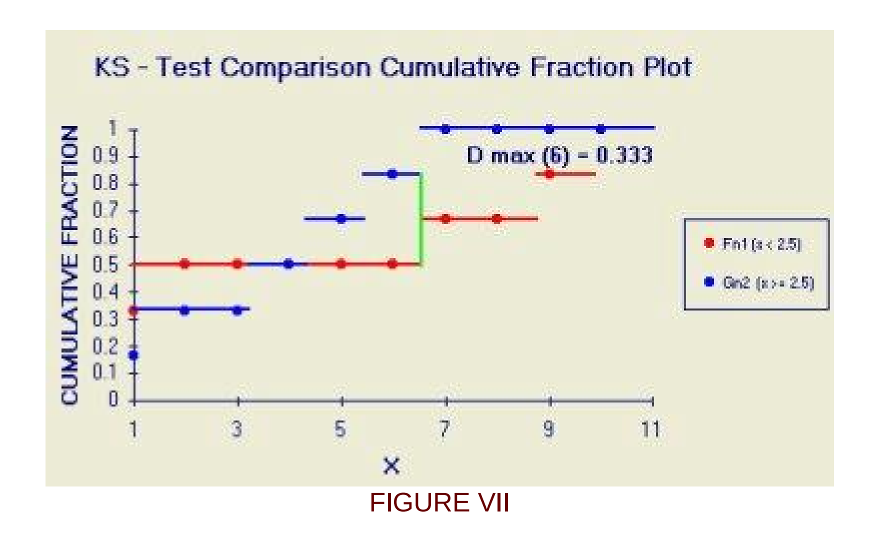


FIGURE VII

Le test n'est pas significatif ; notez qu'un regroupement des catégories comme le pratique l'auteur donne une valeur du $\chi 2$ qui est NS.

13. Test de McNemar

Il s'agit

d'un test qui traite une variable dichotomique [réussite / échec, favorable/défavorable, etc.] mesurée sur les mêmes sujets, à deux moments [avant / après apprentissage, condition 1/ condition 2] notés $\mathbf{t_1}$ et $\mathbf{t_2}$. Ce test se rapproche donc de celui de Wald Wolfowitz et du Sign Test ; il en est congénère. Voici un extrait - du site http://www2.chass.ncsu.edu/garson/pa765/index.shtml - qui permettra de s e faire une exacte idée de l'importance de ce test :

« *The McNemar test assesses the significance of the difference between two dependent samples when the variable of interest is a*

dichotomy. It is used primarily in before-after studies to test for an experimental effect. Also for two dependent samples, the marginal homogeneity test is an extension of the McNemar test for multinomial variables (more than two categorical values). The sign test and the more powerful Wilcoxon signed-ranks test are for two dependent samples when the variable of interest is continuous. [...] *McNemar's test is sometimes called* ***McNemar's test of symmetry*** *or* ***McNemar symmetry chi-square*** *because it, and the marginal homogeneity test which extends it beyond dichotomous data, apply to square tables in which the diagonal reflects subjects who did not change between the* ***before****and****after*** *samples (or matched pair samples). The test of symmetry tests whether the counts in cells above the diagonal differ from counts below the diagonal. If the two counts differ significantly, this reflects change between the samples, such as change due to an experimental effect between the before and after samples.* [...] ***The marginal homogeneity test*** *is similar to the McNemar test, extending it to the case where the variable of interest assumes more than two nominal values. While it may be used with ordinal data, the sign test is preferred.*»

Pour l'utilisation du test d'homogénéité, nous renvoyons donc le lecteur au Sign Test. Sous Excel, la computation de ce test ne pose aucun problème : il faut disposer en deux colonnes contigues les séries d'observations - de taille égale - en deux catégories : par exemple **1** et **0**, et ce avant et après. Voilà un exemple généré avec des nombres aléatoires :

A	1	3	1	3	3	3	3	1	3	1	3	3	1	1	3	3
B	1	3	1	1	3	1	3	3	3	3	1	1	3	1	3	3

	+	**-**
+	6	3
-	4	3

Mc Nemar test

n1,n2 A : 4, B : 3
exact p 0.5
chi2 0
p 1

L'approximation par la loi normale est permise à partir de n1 + n2 >= 25. Deux formules permettent d'obtenir soit la probabilité exacte, soit l'approximation par le chi carré à 1 ddl.

$$p(X \leq r) = \sum_{i=0}^{r} \binom{n_1 + n_2}{i} (0.5)^{n_1 + n_2} \qquad (50)$$

où r = MIN (**n_1,n_2**), donc ici r = 3. La **p** obtenue doit être multipliée par 2 dans le cas d'un test bilatéral.

$$\chi_c^2 = \frac{\left(\left|n_1 - n_2\right| - 1\right)^2}{n_1 + n_2} \qquad (51)$$

Pour appeler le test, il suffit de lancer la boîte de dialogie suivante à partir du menu général :

Mc Nemar test
OK
Annuler
Enter values for n1 and n2
n1 5 n2 20
set of data
labels (1st row) $B$4:$C$4
Input data $B$5:$C$20
1st discrete variable's value 1
2nd discrete variable's value 0
Output data $E$5
© AstroResearch june 2003

Deux possibilités sont offertes, selon que l'utilisateur possède déjà les données à traiter (ici, par exemple, **5** et **20**, où que les données soient brutes ; notez que la macro donne par défaut les valeurs discrètes **1** et **0** par défaut. L'entrée des données se fait comme d'habitude en sélectionnant les plages de cellules voulues. Ici, au cas où l'option **1** serait choisie, le résultat serait :

Mc Nemar test

n1,n2 5, 20
exact p 0.002
chi2 7.84
p 0.005

14. Test de Jonckheere - Terpstra

Alors que le test de Kruskal Wallis - ANOVA non paramétrique - permet de mesurer une différence entre plusieurs moyennes, sans préjuger de leur ordre, le test de J - T permet de mesurer un ordonnancement de différentes moyennes, pourvu que ces échantillons soient indépendants, ce qui différencie la statistique de Jonckheere de celle de Page : voilà donc un test à utiliser pour évaluer, par exemple, un effet dose. Il est donc congénère de la statistique de Page qu'évoque Tomassone à la page 243 de sa ***Biométrie***. Curieusement, le test de J - T n'est pas cité. Il s'agit ici de voir s'il existe une structure sous les traitements. Mais, à la différence de la statistique de Page, ce test ne s'emploie que lorsqu'on est porté à considérer que les échantillons dont l'on dispose sont indépendants.

Il n'a pas été aisé de trouver une documentation substancielle sur ce test. J'ai pu néanmoins sélectionner les sources suivantes :

- capu18.pdf - Testing for trends [School of Psychology University of Nottingham] -
- http://evolution.unibe.ch/teaching/ExpDesign/Exp.Designstats/Kapitel_54_files -
- http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap28/index.htm
- http://software.biostat.washington.edu/%7Erossini/courses/intro-nonpar/text/Computing_the_Jonckheere_Terpstra_Test_.html

Nous donnons ici - en italien mais parfaitement compréhensible pour des oreilles latines - un extrait de la première référence citée :

« *Nel caso di k campioni indipendenti, come nell'analisi della varianza ad un criterio di classificazione,*
quando si suppone che essi siano ordinati secondo il valore delle loro mediane (non importa se in
modo crescente o decrescente), con il test di Jonckheere è possibile verificare l'ipotesi se i vari
campioni o gruppi abbiano tendenze centrali in accordo con la sequenza fissata a priori. E' chiamato anche test di Jonckheere-

Terpstra o delle alternative ordinate (ordered alternatives), in quanto proposto quasi contemporaneamente ed in modo indipendente da T. J. Terpstra nel 1952 (nell'articolo ***The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking*** *pubblicato su Indagationes Mathematicae Vol. 14, pp. 327-333) e da A. R. Jonckheere nel 1954 (****con un articolo intitolato A distribution-free k-sample test against ordered alternatives*** *pubblicato su Biometrika vol. 41, pp. 133-145).* »

Le test a donc été « inventé » d'abord par T.J. Terpstra. L'idée de base est que les moyennes de k échantillons ne diffèrent pas entre elles, considérées du point de vue de leur rang. L'hypothèse nulle $\mathbf{H_0}$ est donc :

$$H_0:\ me_1 = me_2 = me_3 = \ldots = me_k$$ (52)

et l'hypothèse alternative s'énonce comme :

$$H_1:\ me_1 \leq me_2 \leq me_3 \leq \ldots \leq me_k$$ (53)

marche à suivre :
- étape 1 : à partir de k échantillons, construire un tableau où les rangs sont envisagés deux à deux, par un test U de Mann Whitney Wilcoxon :

« *Compute the k(k-1)/2 Mann-Whitney counts U_{ij} , comparing group i with group j , for i < j . The Mann-Whitney count for groups 3 and 4 (for example), are the sum of the number of pairs (X,Y) with X* in group 3, *Y in group 4, and X<Y , i.e. if $X_{i,j}$ is the jth item in the ith group, then [...]* »

La formule générale à employer est, dans le cas présent, si l'on considère le groupe $\mathbf{U_{3,4}}$:

$$U_{3,4} = \sum_{i=1}^{n_3} \sum_{j=1}^{n_4} \phi(X_{3,i}, X_{4,j})$$ (54)

- étape 2 : faire la somme des $\mathbf{U_{i,j}}$:

$$J = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} U_{i,j}$$ (55)

- étape 3 : consulter une table appropriée pour voir si J est supérieur à la valeur critique. Cette table peut être trouvée dans le document pdf en italien - pp. 1447-1449 ; elle est intégrée à la feuille macro et les valeurs critiques sont indiquées au risque 5%, 1% et 0.5%. Pour de grands échantillons - qui correspondent en fait à la taille de ceux pour lesquels les valeurs critiques ne sont pas disponibles, on peut utiliser l'approximation suivante :

$$J^* = \frac{J - E_{H_0}[J]}{\sqrt{\mathrm{Var}_{H_0}(J)}} = \frac{J - \left\{\frac{N^2 - \sum_{j=1}^{k} n_j^2}{4}\right\}}{\sqrt{\frac{N^2(2N+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)}{72}}} \quad (56)$$

Nous indiquons systématiquement la valeur approchée J* à des fins didactiques. Dans (56), $\mathbf{n_j}$ correspond au nombre d'observations du groupe **j**. Voyons cela sur un premier exemple :

A	99	114	116	127	146
B	111	125	143	148	157
C	133	139	149	160	184

	U 1, j	**U 2, j**
U i, 2	18	
U i, 3	23	18

Ce tableau donne la valeurs des rangs Ui,j comparés deux à deux (cf. formule 33). Sous Excel, le résultat apparaît ainsi :

Jonckheere - Terpstra test

N 15
levels 3
$\mathbf{n_i}$ 5 | 5 | 5
$\mathbf{m_i}$ 120.4 | 136.8 | 153
J [x | µ | ±] 59 | 37.5 | 9.46
J* 2.272
p 0.024
critical exact value for J [5 | 1 | 0.5] % 54 | 59 | 62

Pour **J**, nous donnons la valeur [**59**], la moyenne et l'écart type. Rappelons que la moyenne μ_J [**37.5**] est :

$$\mu_J = (N^2 - \sum_{i=1}^{k} n_i^2)/4 \qquad (57)$$

et que la variance σ^2_j [**9.46**] est :

$$\sigma_J^2 = \frac{N^2 \cdot (2N+3) - \sum_{i=1}^{k}\left(n_i^2 \cdot (2n_i+3)\right)}{72} \qquad (58)$$

La valeur de J [**59**] est égale à la valeur critique au risque 1%.
Voyons un autre exemple :

D	12	15	18	20	38	47	48	51	90	108	
C	28	30	38	48	60	66	70	71			
B	31	36	39	44	54	57	63	77	87	123	124
A	35	40	52	67	78	83	88	101	119		

tableau des $\mathbf{U_{i,j}}$

	U 1, j	**U 2, j**	**U 3, j**
U i, 2	51		
U i, 3	78	55	
U i, 4	68	55	58

Jonckheere - Terpstra test

N 38
levels 4
$\mathbf{n_i}$ 10 | 8 | 11 | 9
$\mathbf{m_i}$ 44.7 | 51.38 | 66.82 | 73.67
J [x | µ | ±] 365 | 269.5 | 38.34
J* 2.491
p 0.006
critical exact value for J [5 | 1 | 0.5] % not available

La taille de l'échantillon dépasse les valeurs de la table.
Le test est significatif. L'examen des moyennes montre donc qu'il existe un « effet dose » [en situation unilatérale]
A noter que la statistique de Kruskal Wallis ne permet pas de mettre en évidence de différence significative :

Kruskal Wallis

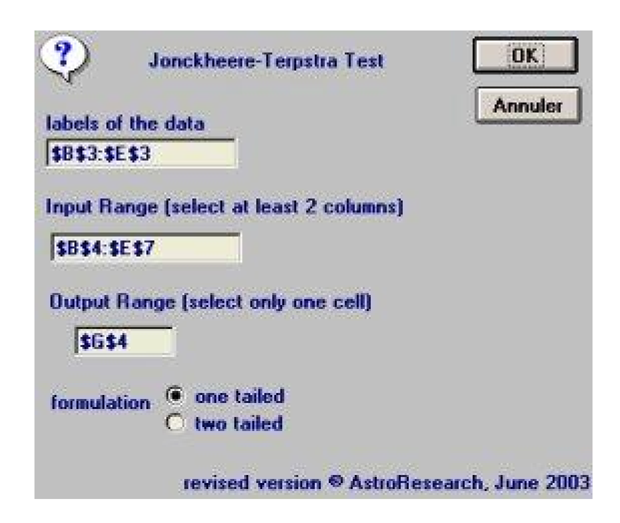
sample size 38
count 10 8 11 9
average 45 51 67 74
rank sum 138 137 240 226
rank average 13.8 17.13 21.82 25.11
ties 4
Q 5.769
Q corr. 5.77
p NS

et que l'ANOVA paramétrique ne permet pas de mettre en évidence de différence entre les doses :

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	5113.23074	3	1704.41025	2.02313695	0.12911837	2.88260082
Within Groups	28643.6114	34	842.459158			
Total	33756.8421	37				

L'ANOVA a été calculée en utilisant l'***Analysis Tool Pack*** d'Excel. La mise en oeuvre du test est semblable aux autres ; à partir de la boîte de dialogue générale, on appelle la boîte suivante :



II. Autres tests statistiques

1)- test de Durbin Watson **è ne figure pas dans l'*Analysis Tool Pack* d'Excel alors qu'un utilitaire de régression est présent...**

Si les résidus obéissent à un « *bruit blanc* », il ne doit pas exister d'autocorrélation
dans une série. On peut alors utiliser entre autre le test de Durbin Watson [test de l • autocorrélation d • ordre 1]. La statistique de Durbin et Watson est une valeur

appartenant à l'intervalle [**0 ; +4**] . Elle est dite normale si elle avoisinne la valeur 2. La valeur **DW** du test de Durbin et Watson peut être anormalement faible ou forte, ce qui peut être l'effet de causes diverses :

- à un processus liant les résidus successifs (dans le cadre des séries chronologiques) ;
- à l'existence d'un changement de structure ;
- au choix d'une spécification inadaptée ;
- à l'omission d'une variable explicative importante ;
  [les variables explicatives sont les vecteurs X ; on les nomme aussi les régresseurs.]

Notons qu'il existe un rapport entre le coefficient de corrélation et la statistique de Durbin et Watson :

- **DW** = 0 si r = +1 (coefficient de corrélation)
- **DW** = 2 si r = 0
- **DW** = 4 si r = -1

Le test de Durbin Watson fait partie de l'ensemble des tests qui figurent dans l'analyse d'une régression linéaire. Voici un exemple :

Sample: 1951 1994
Included observations: 44

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	66.02855	7.314568	9.026992	0.0000
PC	-1.869736	0.573248	-3.261654	0.0022

R-squared	0.202103	Mean dependent var	43.37500
Adjusted R-squared	0.183106	S.D. dependent var	16.83854
S.E. of regression	15.21904	Akaike info criterion	8.327361
Sum squared resid	9728.008	Schwarz criterion	8.408460
Log likelihood	-181.2019	F-statistic	10.63839
Durbin-Watson stat	0.104990	Prob(F-statistic)	0.002203

FIGURE VIII
(*un exemple de résumé d'analyse de régression linéaire*)

Ce test permet, comme d'autres, de mesurer l'importance des résidus et de vérifier, comme tel, les hypothése de la régression : linéarité, normalité, indépendance, variance constante (homoscédasticité). Comme l • espérance des résidus est nulle, on peut écrire :

$$DW = \frac{\sum_{t=2}^{n}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{n}\hat{e}_t^2} \qquad (59)$$

qui définit le test DW. On peut montrer que **DW # 2 - 2.r** où **r** vaut

$$r = \frac{\sum_{t=2}^{n} \hat{e}_{t-1}\hat{e}_t}{\sum_{t=2}^{n} \hat{e}_{t-1}^2} \qquad (60)$$

« *L'indépendance des résidus est une hypothèse fondamentale à considérer car c'est une condition indispensable dans de nombreux tests. Or, si plusieurs mesures sont réalisées sur un même individu, les résidus du modèle déterministe forment une série chronologique et il est intéressant de tester si c'est un bruit blanc ou non. La statistique de Durbin-Watson ou le test de portemanteau (Seber et Wild, 1989, p.322) permettent de tester l'hypothèse que les coefficients de corrélation entre observations successives sont nuls si les observations sont réalisées à intervalles de temps réguliers*. » [F. Husson]

exemple : tiré de Henri Theil, *Principles of Econometrics*, 1971, Wiley, p. 102

[cité in **SHAZAM User's Reference Manual**. A listing of the data set (*filename*: `THEIL.txt`)].

« *This example uses the Theil textile data set. The SHAZAM commands (filename: `DW.SHA`) below first estimate an equation with `PRICE` as the explanatory variable. But economic theory suggests that `INCOME` is an important variable in a demand equation. A statistical result is that if important variables are omitted from the regression then the OLS estimator is biased. The second OLS regression is the preferred model specification that includes both `PRICE` and `INCOME` as explanatory variables.* » [http://shazam.econ.ubc.ca/intro/index.html]

The data is time series for the period 1923 to 1939 (17 observations) for the consumption of textiles in the Netherlands. The variables are:

1. Year
2. Volume of textile consumption per capita (base 1925=100)
3. Real Income per capita (base 1925=100)
4. Relative price of textiles (base 1925=100)

YEAR	CONSUME	INCOME	PRICE
1923	99.2	96.7	101.0
1924	99.0	98.1	100.1
1925	100.0	100.0	100.0

1926	111.6	104.9	90.6
1927	122.2	104.9	86.5
1928	117.6	109.5	89.7
1929	121.1	110.8	90.6
1930	136.0	112.3	82.8
1931	154.2	109.3	70.1
1932	153.6	105.3	65.4
1933	158.5	101.7	61.3
1934	140.6	95.4	62.5
1935	136.2	96.4	63.6
1936	168.0	97.6	52.6
1937	154.3	102.4	59.7
1938	149.0	101.6	59.5
1939	165.5	103.8	61.3

Voici la marche à suivre pour exploiter ces données sous Excel [de la version 4.0 à la version XP] :

a)- sélectionnez les quatre colonnes dans votre browser [ici Mozilla !] ; il faut ensuite les coller où bon vous semble dans une feuille Excel. Vous aurez alors des valeurs dans une seule cellule, séparées par un certain nombre d'espaces [" "] : il s'agit de données non tabulées. Une macro va permettre de restituer à chaque colonne sa propre colonne dans Excel, c'est-à-dire de la tabuler. [la macro a été rentrée dans la colonne B et **space** est son nom, dans la cellule B1].

space

```
=POUR("y";1;40)
=POSER.VALEUR(B4;NBCAR(CELLULE.ACTIVE()))
0
=POUR("x";1;B4)
=SI(B4=0;ATTEINDRE(B30);)
=CHERCHE(" ";CELLULE.ACTIVE();1)
```

```
=POSER.VALEUR(B9;B7)
6
=GAUCHE(CELLULE.ACTIVE();B9-1)
=POSER.VALEUR(B12;B10)
103.8
=SELECTIONNER("rc"&COLONNE(CELLULE.ACTIVE())+x&"")
=FORMULE(B12)
=SELECTIONNER("rc"&COLONNE(CELLULE.ACTIVE())-x&"")
=REMPLACER(CELLULE.ACTIVE();1;B9;"")
=FORMULE(B16)
=CHERCHE(" ";CELLULE.ACTIVE();1)
=POSER.VALEUR(B20;B18)

=SI(ESTNUM(B20);ATTEINDRE(B26);POSER.VALEUR(B22;CELLULE.ACTIVE()))
61.3
=SELECTIONNER("rc"&COLONNE(CELLULE.ACTIVE())+x+1&"")
=FORMULE(B22)
=SI(ESTNUM(B20);;ATTEINDRE(B27))
=SUIVANT()
=SELECTIONNER("r"&y+1&"c"&COLONNE(CELLULE.ACTIVE())-SI(B4=0;-1;x)-1&"")
=SI(NBCAR(CELLULE.ACTIVE())=0;ATTEINDRE(B30);ATTEINDRE(B31))
=EDITION.SUPPRIMER(2)
=SUIVANT()
=RETOUR()
```

Cette macro ne marche que s'il y a un seul " ". En cas de plusieurs "   " [ici : 3], il est plus simple d'employer d'abord la commande REMPLACER du menu où vous demandez à remplacer à chaque fois 2 " " par 1 seul " ". La macro marchera correctement à partir du moment où les valeurs numériques ne seront plus séparées que par un seul " ".

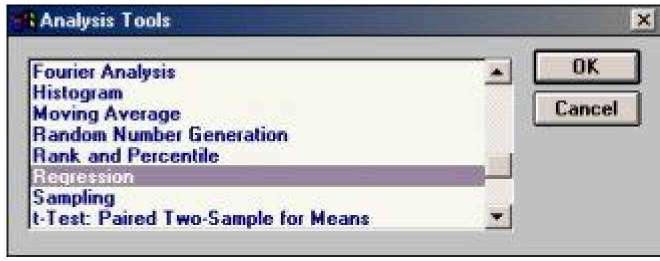
b)- on aboutit au tableau de données suivant, après avoir disposé les étiquettes :

YEAR	CONSUME	INCOME	PRICE
1923	99.2	96.7	101
1924	99	98.1	100.1
1925	100	100	100
1926	111.6	104.9	90.6
1927	122.2	104.9	86.5
1928	117.6	109.5	89.7
1929	121.1	110.8	90.6
1930	136	112.3	82.8
1931	154.2	109.3	70.1
1932	153.6	105.3	65.4
1933	158.5	101.7	61.3
1934	140.6	95.4	62.5
1935	136.2	96.4	63.6
1936	168	97.6	52.6
1937	154.3	102.4	59.7
1938	149	101.6	59.5
1939	165.5	103.8	61.3

TABLEAU XXI

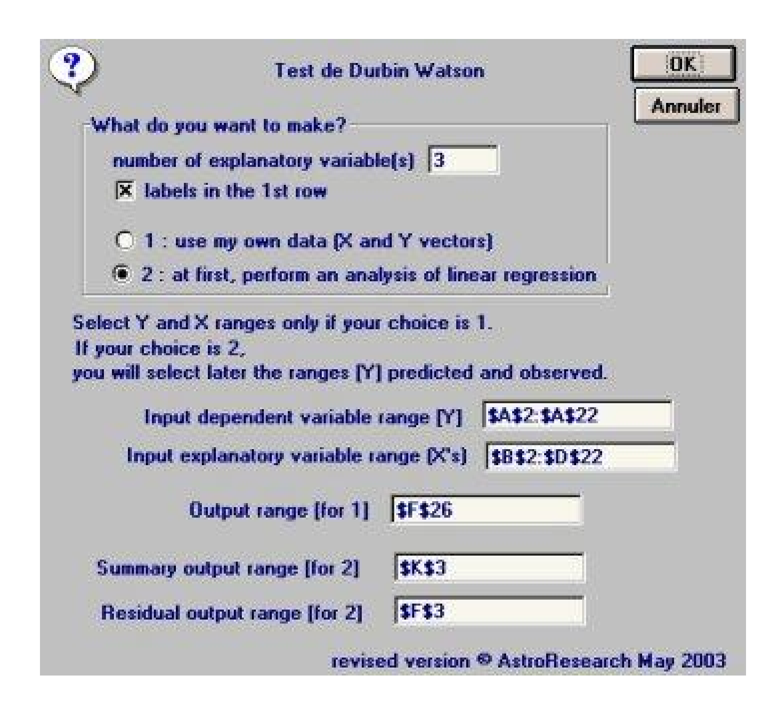
(*tableau de données tabulées*)

c)- il faut lancer la procédure de régression linéaire. Le plus simple consiste à utiliser d'abord les ressources de l'***Analysis Tool Pack*** d'Excel et de sélectionner l'outil régression :



(outil régression)

L'astuce va alors consister à substituer une boîte de dialogue propre à notre macro qui va permettre de « piloter » l'outil régression.



L'accent est mis sur le fait que l'utilisateur peut employer soit ses propres données déjà calculées [predicted Y], soit qu'il va réaliser une première analyse. Posons que nous soyons dans ce deuxième cas de figure. La figure suivante montre les équivalences de zones de saisie entre la boîte de dialogue du DW et celle de l'outil régression. En effet, pour l'utilisateur, le fonctionnement de l'outil régression sera « transparent » puisque la boîte de dialogue propre à cet outil n'apparaîtra pas. Si nous reprenons les données de Shazam, nous obtenons d'abord, en sélectionnant 2 [at first, perform an analysis of linear regression], les données suivantes :

Regression Statistics

Multiple R	0.9753367
R Square	0.95128167

Adjusted R Square 0.94432191
Standard Error 5.56335574
Observations 17

Analysis of Variance

df Sum of Squares Mean Square F
Significance F
Regression 2 8460.93643 4230.46822 136.683086 6.514E-10
Residual 14 433.312979 30.950927
Total 16 8894.24941

Coefficients Standard Error t Statistic P-value Lower 95% Upper 95%

Intercept 130.706587 27.0942926 4.8241373 0.00018691 72.5950577 188.818117
INCOME 1.06170963 0.26667397 3.98130214 0.00107341 0.48975035 1.63366891
PRICE -1.38298546 0.08381426 -16.5005981 1.8141E-11 -1.56274933 -1.20322158

4)- le test de Durbin-Watson

A partir des données de la figure IV où l'on prend **CONSUME** comme variable dépendante et **INCOME** et**PRICE** comme explicatives [régresseurs], on aboutit aux données suivantes :

CONSUME	Predicted Y
99.2	93.6923774
99	96.4234578
100	98.5790046
111.6	116.781445
122.2	122.451685
117.6	122.909996
121.1	123.045532
136	135.425383
154.2	149.804169
153.6	152.057362
158.5	153.905448
140.6	145.557095
136.2	145.097521
168	161.584412
154.3	156.861422
149	156.288651
165.5	156.135038

Le vecteur **CONSUME** reste celui saisi dans la colonne A, tel qu'indiqué à la figure VI ; il faut par contre modifier la saisie des variables X pour sélectionner, en lieu et place,

le **PREDICTED Y** fourni par l'***Analysis Tool Pack*** d'Excel [outil régression]. Il y a une différence entre Excel 4.0 - 5.0 et les versions ultérieures : c'est que les valeurs prédites de Y, les résidus et les résidus standardisés sont saisis au-dessous de l'analyse elle-même alors qu'ils étaient séparés dans les premières versions. Aussi n'est-il point besoin, dans les versions postérieures à Excel 6.0 de saisir la cellule de « *Residual output range* ».

Durbin Watson test	
N	17
m	2
DW	2.019
r	-0.009
critical value	1.02
critical value	1.54
p	NS
autocorrelati	0

TABLEAU XXII
(*test de Durbin Watson*)

Le tableau XII est obtenu après avoir sélectionné 1 dans la boîte de dialogue de la figure VI. est le nombre d'observations, **m** le nombre de variables dites explicatives [ici 2]. r est le coefficient de corrélation qui suit l'approximation :

$$\rho = -\left(\frac{DW - 2}{2}\right) \quad (61)$$

Le sens de la valeur de **NDW** [2.019] peut être comrpis et illustré d'après la figure suivante :

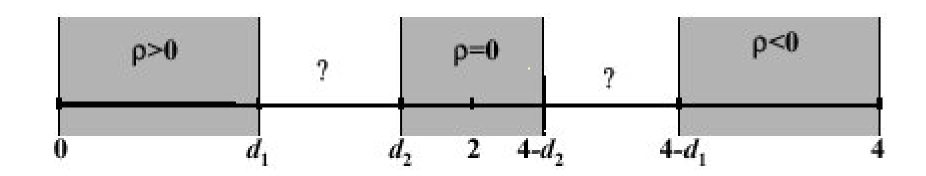


FIGURE IX
(*le sens de l'interprétation du test de DW*)

Selon la position de **DW** on peut conclure :

Regions of Acceptance and Rejection of the Null Hypothesis				
Zero to d_l	d_l to d_u	d_u to $(4 - d_u)$	$(4-d_u)$ to $(4-d_l)$	$(4-d_l)$ to 4
Reject Null H_o: **Positive** Autocorrelation	Neither accept or reject	Accept the Null Hypothesis	Neither accept or reject	Reject Null H_o: **Negative** Autocorrelation

FIGURE X

Dans notre cas de figure, **DW** = 2.019 et compris entre $\mathbf{d_2}$ [1.02] et **4 -** $\mathbf{d_2}$ [2.46 = 4 - 1.54]. On accepte donc H_0 [autocorrelation = 0].

Les tables pour interpréter la valeur du **DW** sont incluses dans la feuille macro et consultables sous forme de fichiers image. On en trouvera en outre une version au format *pdf*
[au risque 5% et au risque 1%].

II. Le logiciel

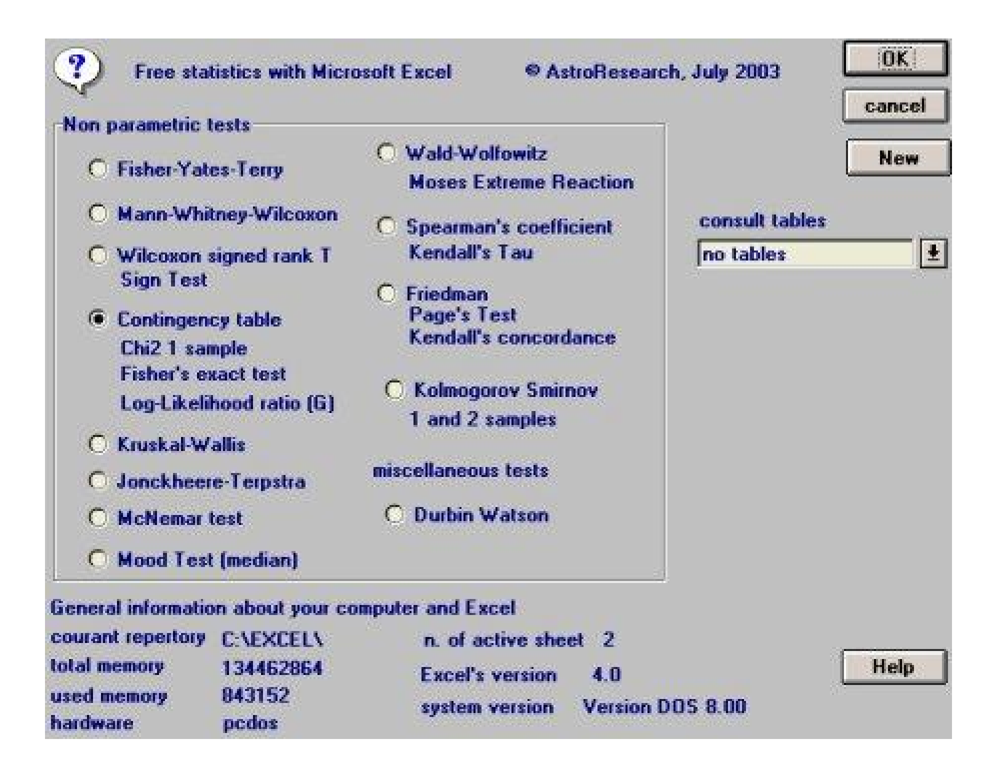
Il se compose d'une feuille macro "**MANNWHIT.XLM**" au format Excel 4.0. Quand on lance **MANNWHIT.XLM**, une feuille de calcul est bâtie :

MOOD2.XLS

	A	B	C	D	E	F	G
1							
2	*Click the button to appear tests or type the sequence ctrl - a*						
3							
4							
5		Primaire	Collège	Terminale	Université (1	Univ. (2)	Univ. (>2)
6		4	2	2	9	2	2
7		3	4	0	4	4	6
8		0	1	4	2	5	
9		7	6	3	3	2	
10		1	3	8			
11		2	0	0			
12		0	2	5			
13		3	5	2			
14		5	1	1			

TABLEAU XXIII

Le bord supérieur gauche de la cellule contient un bouton sur lequel on clique pour faire apparaître le menu. On peut aussi se servir de la séquence de touches **ctrl-a**. On saisit ensuite les données dans des colonnes adjacentes. Voici le menu général sous la forme d'une boîte de dialogue :



Il y a actuellement 19 tests disponibles, y compris les tables pour les effectifs réduits [23 tables]. Pour accéder au test exact de Fisher, il suffit de sélectionner la table de contingence ; si le tableau est de taille **2** x **2** et que les effectifs théoriques soient inférieurs à 3 ou à une valeur à fixer, le test de Fisher sera activé. Mais vous pouvez aussi forcer son utilisation par une case à cocher. Pour accéder au test voulu, il suffit de sélectionner la case d'option. Notez que certains tests ont été regroupés [ex : Friedman, Page et W de Kendall, cf. supra] Pour consultez les tables, il suffit de sélectionner la liste déroulante et d'afficher un autre niveau que « ***no tables*** » visible ci-dessus, qui constitue l'option par défaut [le menu est alors désactivé ; n'oubliez pas de sélectionner à nouveau « ***no tables*** » pour pouvoir accéder à l'ensemble des tests]. Quand le test est choisi, la boîte de dialogue spécifique au test apparaît, comme sur la figure suivante :

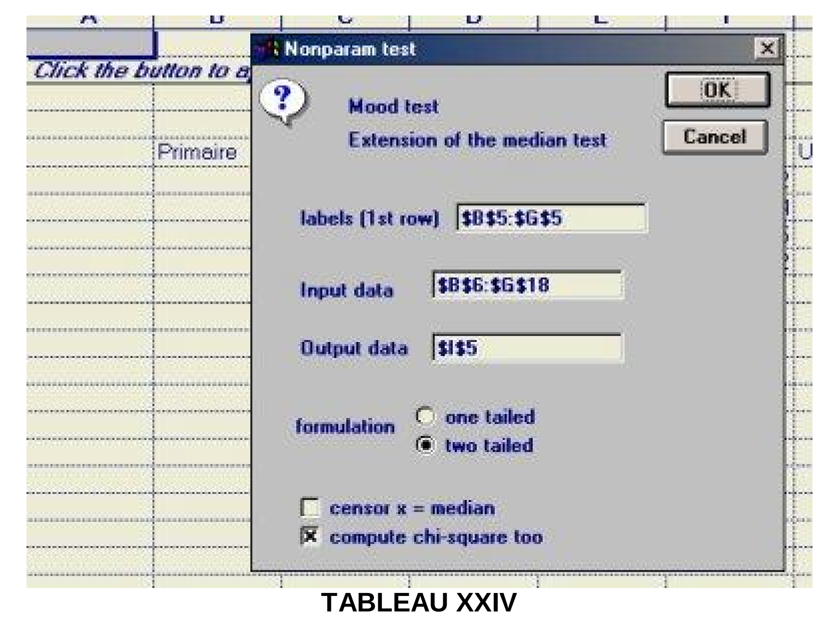


TABLEAU XXIV

(cf Extension of the median test , Biométrie, p. 206)

Comme on le voit, la boîte de dialogue s'inspire totalement de celles existant dans ***l'Analysis Tool Pack***. Dans le cas présent, 2 colonnes contigues sont sélectionnées [A5:B12] et le résultat apparaîtra dans la cellule D5, sélectionnée par l'utilisateur. On appuie ensuite sur le bouton OK et un résultat analogue à celui du tableau III apparaît.

Pour l'instant, seuls 19 tests figurent. Compte tenu que la programmation me prend un temps important, pendant lequel je ne puis faire progresser mes autres travaux, le lecteur comprendra que je n'étofferais davantage ces options que si la demande s'en fait réellement sentir. Qu'il sache simplement que ce logiciel est évidemment « freeware », puisque développé sous Microsoft® Excel 4.0 pour Windows!" 3.1 [n° de licence 00-065-0400-L5433225]. Par comparaison, de nombreux logiciels intègrent ces tests mais coûtent plus de 500 ¬... Et les logiciels développés librement ne me semblent pas avoir toute la souplesse et le confort d'utilisation qui est propre à l'environnement d'Excel.

Bibliographie

a. générale

1. ARMITAGE P, BERRY P et BLACKWELL G ., 1994, Statistical Methods in Medical Research, Ed. Scientific Publications, Oxford.
2. COHEN L., HOLLIDAY L . et M., 1983, Statistics for Social Scientists, Ed. Harper and Row.
3. GARDNER M. et ALTMAN D.G., 1989, Statistics with Confidence, Ed. British Journal Publications, Londres.
4. ROSENTHAL R. et ROSNOW R.L.,1991, Essentials of Behavioral

Research, Ed. McGraw-Hill, Series in Psychology.
5. SCHWARTZ D.,1970, Méthodes statistiques à l'usage des médecins et des biologistes, Éd. Flammarion
6. TOMASSONE R.,DERVIN C.et MASSON J.P,1992, Biométrie ; modélisation de phénomènes biologiques, Éd. Masson. [un must, écrit dans un style aussi clair qu'il est possible et à échelle « humaine » ; en plus, nous avons droit à quelques notes de ***Don Giovanni*** en préface : voilà une bonne entrée en matière !]
7. ZAR J.H., 1996, Biostatistical Analysis, Prentice Hall International Editions.
8. KYD Ch.W., KINATA C., Les Macros d'Excel, DunodTECH, Paris, 1992
9. FISHER R.A., 1946.*Statistical methods for research workers*, Olivier & Boyd, London. traduction française aux Presses Universitaires.
10. Bertrand M. & Charles B. (1987) - *Initiation à la statistique et aux probabilités* - Eyrolles.
11. Centre d'Enseignants et de Recherche de Statistique Appliquée (1986) - *Aide-mémoire pratique des techniques statistiques pour ingénieurs et techniciens supérieurs* - Ceresta.
12. Falissard B. (1996) - *Comprendre et utiliser les statistiques dans les sciences de la vie* -Masson.
13. Abdi, H. (1987). Introduction au traitement statistique des données expérimentales. Grenoble: Presses Universitaires de Grenoble.
14. Baron, J., & Li, Y. (2000, 31 decembre). Notes on the use of R for psychology experiments and questionnaires. (disponible à http://www.psych.upenn.edu/ baron/rpsych.htm)
15. Bradley, James (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall.

b. tests non paramétriques

1. CAPERAA Philippe & VAN CUTSEM Bernard, 1988. ***Méthodes et modèles en statistique non paramétrique. Exposé fondamental***. Presses Université Laval, Dunod, 357 pp.
2. SIEGEL Sidney, 1956. ***Non parametric statistics for the behavioral sciences*** , McGraw Hill, 312 pp.
3. SPRENT P. 1992. ***Pratique des statistiques non parmétriques***. INRA Editions.
4. Kruskal, W.H. and Wallis, W.A. (1952) ***Use of ranks in one-criterion variance analysis***. Jour. Am. Stat. Assoc. 47. 583-634
5. Kruskal, W.H. and Wallis, W.A. (1952) ***Errata to Use of ranks in one-criterion variance analysis***. Jour. Am. Stat. Assoc. 48. 907-911.
6. Wallace, D.L. (1959). ***Simplified beta-approximations to the Kruskal-Wallis H test***. Jour. Am. Stat. Assoc. 54. 225-230.
7. Griffiths, D. (1980). ***A Pragmatic Approach to Spearman's Rank Correlation Coefficient***. Teaching Statistics 2, pp. 10?13.
8. Kruskal, W. *(1958).****Ordinal Measures of Association***. Journal of the American Statistical Association *53,* pp. 814?861.
9. Leach, C. (1979). ***Introduction to Statistics: A Nonparametric Approach for the Social Sciences***. Wiley.
10. Wilkie, D. (1980). ***Pictorial Representation of Kendall's, Rank Correlation Coefficient***. Teaching Statistics 2, pp. 76-78.
11. S. Siegel and N. J. Castellan, ***Nonparametric Statistics***, 2nd ed., 1988
12. Moses, L. E. (1952).***A two-sample test***. Psychometrika, 17, 234 • 247

13. Moses, L. E. (1963). ***Rank tests of dispersion***. Annals of Mathematical Statistics, 34, 973.
14. Dineen, L. C., and Blakesley, B. C. 1973. ***Algorithm AS 62: Generator for the sampling distribution of the Mann-Whitney U statistic***. Applied Statistics, 22: 269 • 273.
15. Lehmann, E. L. 1985. ***Nonparametrics: Statistical Methods Based on Ranks***. San Francisco: McGraw Hill.
16. Smirnov, N. V. 1948. ***Table for estimating the goodness of fit of empirical distributions***. Annals of Mathematical Statistics, 19: 279 • 281.

bibliographie complémentaire [adaptée et tirée de : Clint W. Coakley, November 1998]

One Sample Preliminary Tests

1. Rank von Neumann Ratio test of independence

Bartels, R. (1982), "The Rank Version of von Neumann's Ratio Test for Randomness," *Journal of the American Statistical Association*,**77**, 40-46.
Gibbons, J. D. and Chakraborti, S. (1992), *Nonparametric Statistical Inference*, New York: Marcel Dekker, pp. 88-90 (textbook treatment).

2. Kolmogorov-Smirnov Goodness-of-Fit test

Kolmogorov, A. N. (1933), "Sulla determinazione empirica di una legge di distribuzione," *Giorn. Dell' Inst. Ital. Degli Attuari*,**4**, 83-91.
Conover, W. J. (1980), *Practical Nonparametric Statistics*, second edition, New-York: John Wiley & Sons, Inc., pp. 344-356 (textbook treatment, applied).
Gibbons, J. D. and Chakraborti, S. (1992), *Nonparametric Statistical Inference*, New York: Marcel Dekker, pp. 104-120 (textbook treatment, theory), p. 487 for Table F (null distribution).

3. Pearson Chi-squared Goodness-of-Fit test

Pearson, K. (1900), "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine*, Series 5, **50**, 157-175.
Daniel, W. W. (1990), *Applied Nonparametric Statistics*, Boston: PWS-Kent, pp. 306-316 (textbook treatment, applied).
Gibbons, J. D. and Chakraborti, S. (1992), *Nonparametric Statistical Inference*, New York: Marcel Dekker, pp. 95-103 (textbook treatment, theory).

4. Shapiro-Wilk test of normality

   Shapiro, S. S. and Wilk, M. B. (1965), "An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591-611
   Conover, W. J. (1980), *Practical Nonparametric Statistics*, second edition, New-York: John Wiley & Sons, Inc., pp. 363-367 for text, pp. 466-470 for Tables A17-A19 (coefficients and null and approximate distributions).

5. Ryan-Joiner version of the Shapiro-Wilk test (Minitab's normplot macro)

   Ryan, T. A. and Joiner, B. L. (1976), "Normal Probability Plots and Tests for Normality," Technical Report, Minitab, Inc. State College PA.

6. Stephens' modification of the K-S test

   Stephens, M. A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association*, **69**, 730-737.

7. Boos test of symmetry

   Boos, D. D. (1982), "A Test for Asymmetry Associated with the Hodges-Lehmann Estimator, " *Journal of the American Statistical Association*, **77**,

   647-651.

8. ELR test of symmetry

   Eubank, R. L., LaRiccia, V. N., and Rosenstein, R. B. (1992), "Testing Symmetry about an Unknown Median via Linear Rank Procedures," *Journal of Nonparametric Statistics*, **1**, 301-311.

9. Triples test of symmetry

   Randles, R. H., Fligner, M. A., Policello, G. E., and Wolfe, D. A. (1980), "An Asymptotically Distribution Free Test for Symmetry Versus Asymmetry," *Journal of the American Statistical Association*, **75**, 168-172.

Location Tests for One Sample or Paired Samples

1. The t test (classical) "Student" (or Gosset, W. S.) (1908), "On the probable error of the mean," *Biometrika*, **6**, 1-25.

2. The sign test

Dixon, W. J. and Mood, A. M. (1946), "The statistical sign test," *Journal of the American Statistical Association*, **41**, 557-566.
Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd. This is credited with one of the first uses of the sign test.
Thompson, W. R. (1936), "On confidence ranges for the median and other expectation distributions for populations of unknown distribution form," *Annals of Mathematical Statistics*, **7**, 122-128. This paper derived the sign based confidence interval for the median.
Hettmansperger, T. P. and Sheather, S. J. (1986), "Confidence intervals based on interpolated order statistics," *Statistics & Probability Letters*, **4**, 75-79. This gives the nonlinear interpolation formula for CI's used by Minitab.

3. The Wilcoxon signed ranks test

Wilcoxon, F. (1945), "Individual comparisons by ranking methods," *Biometrics*, **1**, 80-83.
Hodges, J. L., Jr. and Lehmann, E. L. (1963), "Estimates of location based on rank tests," *Annals of Mathematical Statistics*, **33**, 482-497. This paper proposed the median of Walsh averages as an estimator of the median.
Daniel, W. W. (1990), *Applied Nonparametric Statistics*, second edition, Boston: PWS-Kent, pp. 496-502 for Table A.3 (null distribution).

4. The (one sample) normal scores test

Fraser, D. A. S. (1957), *Nonparametric Methods in Statistics*, New York: John Wiley & Sons, Inc.

Two Sample Location Tests

1. The two sample Welch t test (also known as the Smith-Welch-Satterthwaite t test)

Welch, B. L. (1937), "The Significance of the Difference Between Two Means When the Population Variances are Unequal," *Biometrika*,**29**, 350-362.
Smith, H. F. (1936), "The Problem of Comparing the Results of Two Experiments With Unequal Errors,"

Journal of the Council for Scientific and Industrial Research, **9**, 211-212.
Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components,"*Biometric Bulletin*, **2**, 110-114.

2. The Mann-Whitney-Wilcoxon test (or Wilcoxon rank sum test)

Wilcoxon, F. (1945), "Individual comparisons by ranking methods," *Biometrics*, **1**, 80-83.
Mann, H. B. and Whitney, D. R. (1947), "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, **18**, 50-60.
Gibbons, J. D. and Chakraborti, S. (1992), *Nonparametric Statistical Inference*, New York: Marcel Dekker, pp. 495-502 for Table J (null distribution).

3. Mood's median test

Mood, A. M. (1950), *Introduction to the Theory of Statistics*, New York: McGraw-Hill. The most frequently cited source for this test.
Westenberg, J. (1948), "Significance Test for Median and Interquartile Range in Samples from Continuous Populations of Any Form," *Akad. Wetensch. Afdeeling Voor de Wis.*, **51**, 252-261. The earliest known source, although the title is misleading with respect to the assumptions.

4. The modified Mathisen test

Hettmansperger, T. P. and McKean, J. W. (1998), *Robust Nonparametric Statistical Methods*, London: Arnold, pp. 105, 131-133.
Mathisen, H. C. (1943), "A method of testing the hypothesis that two samples are from the same population," *Annals of Mathematical Statistics*, **14**, 188-194. This paper proposed the original Mathisen test, known as the control median test.

5. The two sample normal scores test

van der Waerden, B. L. (1952/1953), "Order tests for the two sample problem and their power," I. Indagationes Mathematicae, **14**, 453-458; II. Indagationes Mathematicae, **15**, 303-310; III. Indagationes Mathematicae, **15**, 311-316. (Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen 55 and 56).

Two Sample Preliminary Tests

1. The Two Sample Kolmogorov-Smirnov Test

   Smirnov, N. V. (1939), "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," (Russian) *Bull. Moscow Univ.*, **2**, 3-16.
   Daniel, W. W. (1990), *Applied Nonparametric Statistics*, second edition, Boston: PWS-Kent, pp. 574-576 for Table A.20 (null distribution).

2. The F test for the ratio of two normal variances (classical)

   Ott, R. L. (1993), *An Introduction to Statistical Methods and Data Analysis*, fourth edition, Belmont CA: Wadsworth, pp. 340-345. Many other books can be cited for this test. I do not know the original source, but it is probably Fisher (1925) (see the references for the sign test).

3. The Moses two sample dispersion test

   Moses, L. E. (1963), "Rank tests of dispersion," *Annals of Mathematical Statistics*, **34**, 973-983.
   Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 93-102 (textbook treatment).

4. The Ansari-Bradley two sample dispersion test

   Ansari, A. R. and Bradley, R. A. (1960), "Rank-sum tests for dispersions," *Annals of Mathematical Statistics*, **31**, 1174-1189.
   Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 83-93 (textbook treatment).

5. The Miller Jackknife two sample dispersion test

   Miller, R. G., Jr. (1968), "Jackknifing variance," *Annals of Mathematical Statistics*, **39**, 567-582.
   Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 103-111 (textbook treatment).

Tests for the One Way Layout (k sample problem)

1. Bartlett's test of homogeneity of variance

Bartlett, M. S. (1937), "Properties of sufficiency and statistical tests," *Proceedings of the Royal Society*, A**160**, 268-282.

2. Levene's test of homogeneity of variance

Levene, H. (1960), "Robust tests for equality of variances," in *Contributions to Probability and Statistics*, Palo Alto CA: Stanford University Press, pp. 278-292.
Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996), *Applied Linear Statistical Models*, Chicago: Irwin, pp. 766-768 (textbook treatment).

3. Welch ANOVA (for unequal variances)

Welch, B. L. (1951), "On the comparison of several mean values: an alternative approach," *Biometrika*, **38**, 330-336.

4. Kruskal-Wallis test

Kruskal, W. H. and Wallis, W. A. (1952), "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, **47**, 583-621.
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 114-119 (textbook treatment).
Kraft, C. H. and van Eeden, C. (1968), A Nonparametric Introduction to Statistics, New York: Macmillan, pp. 238-262 for Table F (null distribution).

5. Multiple comparisons in the one way layout

Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 124-132 (textbook treatment of exact, Scheffe, and Tukey approaches).
Dunn, O. J. (1964), "Multiple comparisons using rank sums," *Technometrics*,**6**, 241-252 (Bonferroni approach).
Miller, R. G., Jr. (1966), *Simultaneous Statistical Inference*, New York: McGraw-Hill. This book covers many classical and nonparametric multiple comparison procedures.
Hollander, M. H. and Wolfe, D. A. (1973),

Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 328-334 for Tables A.9 - A.12.

6. Jonckheere-Terpstra test for ordered alternatives

Jonckheere, A. R. (1954), "A distribution-free k-sample test against ordered alternatives," *Biometrika*, **41**, 133-145.
Terpstra, T. J. (1952), "The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking," *Indagationes Math.*, **14**, 327-333.
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 120-123 (textbook treatment).
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 311-327 for Table A.8 (null distribution).

Tests for the Two Way Layout

1. Friedman test

Friedman, M. (1937), "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, **32**, 675-701.
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 138-146 (textbook treatment).
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 366-371 for Table A.15 (null distribution).

2. Multiple comparisons in the two way layout

Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp.151-158 (textbook treatment of exact and Tukey approaches).
Miller, R. G., Jr. (1966), *Simultaneous Statistical Inference*, New York: McGraw-Hill.
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 373-382 for Tables A.17-A.19.

3. Page test for ordered alternatives

Page, E. B. (1963), "Ordered hypotheses for multiple treatments: a significance test for linear ranks," *Journal of the American Statistical Association*, **58**, 216-230.
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 147-150 (textbook treatment).
Daniel, W. W. (1990), *Applied Nonparametric Statistics*, Boston: PWS-Kent, pp. 570 for Table A.17 (null distribution).

4. Durbin test for incomplete block designs

Durbin, J. (1951), "Incomplete blocks in ranking experiments," *British Journal of Statistical Psychology*, **4**, 85-90.

Procedures for Correlation and Simple Regression

1. Pearson correlation (classical)

Galton, F. (1888), "Co-relations and their measurement, chiefly from anthropological data," *Proceedings of the Royal Society of London*, **45**, 135-145.

2. Spearman correlation coefficient

Spearman, C. (1904), "The proof and measurement of association between two things," *American Journal of Psychology*, **15**, 72-101.
Siegel, S. and Castellan, N. J., Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, second edition, New York: McGraw-Hill, pp. 235-244 (textbook treatment, applied).

3. Kendall’s tau

Kendall, M. G. (1938), "A new measure of rank correlation," *Biometrika*,**30**, 81-93.
Siegel, S. and Castellan, N. J., Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, second edition, New York: McGraw-Hill, pp. 245-254 (textbook treatment, applied).
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 384-393 for Table A.21 (null distribution).

4. Theil-Sen simple linear regression procedures

Theil, H. (1950), "A rank-invariant measure of linear and polynomial regression analysis," I. *Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen,* A**53**, 386-392; II. *Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen,* A**53**, 521-525; III.
Proceedings of Koninklijke Nederlandse Akademie vanWetenschappen, A**53**, 1397-1412.
Sen, P. K. (1968), "Estimates of the regression coefficient based on Kendall's tau," *Journal of the American Statistical Association*,**63**, 1379-1389.
Sen generalized Theil's procedure to the case where there are ties among the regressor values.
Sprent, P. (1993), *Applied Nonparametric Statistical Methods*, second edition, London: Chapman and Hall, pp. 188-202 (textbook treatment, applied).
Hollander, M. H. and Wolfe, D. A. (1973), Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc., pp. 384-393 for Table A.21 (null distribution).

5. Kernel regression

Nadaraya, E. A. (1964), "On estimating regression," *Theory of Probability and Its Applications,* **9**, 141-142.
Watson, G. S. (1964), "Smooth regression analysis," *Sankhya*, Series A, **26**, 359-372.
Ryan, T. P. (1997), *Modern Regression Methods*, New York: John Wiley & Sons, Inc., Chapter 10 (textbook treatment).

6. Local linear regression

Stone, C. J. (1977), "Consistent nonparametric regression," *Annals of Statistics*, **5**, 595-645.
Cleveland, W. S. (1979), "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association,* **74**, 829-836.
Ryan, T. P. (1997), *Modern Regression Methods*, New York: John Wiley &Sons, Inc., Chapter 10 (textbook treatment).

Rank-based Analysis of Linear Models

McKean, J. W. and Hettmansperger, T. P. (1976), "Tests of hypotheses based on ranks in the general linear model," *Communications in Statistics – Theory and Methods,* A**5**, 693-709.
Hettmansperger, T. P. and McKean, J. W. (1977), "A robust alternative based on ranks to least squares in analyzing linear models," *Technometrics*,**19**, 275-284.

Hettmansperger, T. P. (1984), *Statistical Inference Based on Ranks*, New York: John Wiley & Sons, Inc., Chapter 5.
Hettmansperger, T. P. and McKean, J. W. (1998), *Robust NonparametricStatistical Methods*, London: Arnold.

c. liens sur internet

Ils sont bien sûr très nombreux mais j'ai pu en sélectionner quelques-uns qui ont le mérite de proposer soit l'intégralité des tests [théorie et méthode], soit des exemples, soit les deux. La plupart des articles ou des cours sont au format pdf [Acrobat Reader] ou au format ps [Ghostscript, nécessitant l'application Ghostscript et surtout le « viewver », Ghostview]. Voici mes choix :

1. http://www.unesco.org/webworld/idams/advguide/Chapt4_2_files/filelist.xml
2. Statistique et probabilité : A. Bar-Hen (d'apres J. Roussel), Université Aix-Marseille II [donne entre autre de nombreuses tables]
3. méthodes statistiques pour l'ingénieur, Olivier Gaudouin, ENSIMAG 2ème année, INP, Grenoble
4. maîtrises STAPS de l'Université Antilles-Guyane, Michel Le-Her
5. Centre d'Enseignants et de Recherche de Statistique Appliquée (1986) - *Aide-mémoire pratique des techniques statistiques pour ingénieurs et techniciens supérieurs* - Ceresta.
6. Tests statistiques, note pédagogique, Christophe Benavent Pr. à l'IAE de Lille
7. DEA Analyse et Modélisation des Systèmes Biologiques Introduction au logiciel S-PLUS© D. Chessel
8. Use of statistical programs for nonparametric tests of small samples often leads to incorrect P values: examples from Animal Behaviour, ROGER MUNDRY & JULIA FISCHER Institut für Verhaltensbiologie, Freie Universität Berlin, ANIMAL BEHAVIOUR, 1998, 56, 256 • 259
9. Simplified Procedure for Implementing Nonparametric Tests in Excel, Robert J. Pavur, University of North Texas, and Kellie B. Keeling, Virginia Tech, RICK HESSE, Feature Editor, Graziadio Graduate School of Business and Management, Pepperdine University
10. BASIC NON-PARAMETRIC STATISTICAL TOOLS, prepared for GCMA 200, Peter M. Quesada, Gregory S. Rash. Examples presented in these notes were obtained from Primer of Biostatistics by Stanton S. Glantz (McGraw Hill Text; ISBN: 0070242682)
11. Nonparametric tests, from Minitab data.
12. Use of Durbin-Watson Test Statistic : exact P values, Gould
13. The Econometric Journal of line, Econometric Software Links Econometrics Journal
14. EViews 4.0 User • s Guide
15. Linear Regression Analysis, Pr. Roy Batchelor City University Business School, London ESCP, Paris
16. Arthur Charpentier, cours de série temporelle, DESS Mathématiques de la décision, et DESS Actuariat
17. http://wsupsy.psy.twsu.edu/charlie/nonparametric_tests.htm
18. Guide to Advanced Data Analysis using IDAMS Software P.S.

NAGPAUL, New Delhi (India)

19. http://www.dsa.unipr.it/soliani/soliani.html. Cf. supra :

MANUALE DI STATISTICA
PER LA RICERCA E LA PROFESSIONE

STATISTICA UNIVARIATA E BIVARIATA
PARAMETRICA E NON-PARAMETRICA
PER LE DISCIPLINE AMBIENTALI E BIOLOGICHE
(edizione febbraio 2003)

Lamberto Soliani
con la collaborazione di
Franco Sartore e Enzo Siri

avec pas moins de 32 chapitres organisés selon des fichiers au format pdf. Ceux traitant des tests non paramétriques sont les chapitres : **caput : 6 - 7 - 8 - 14 - 18**.

d. logiciels

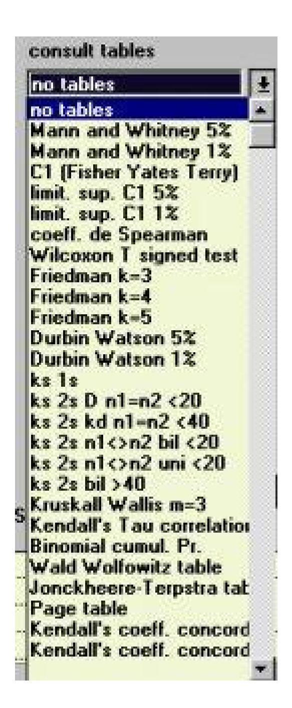
1. Logiciels spécifiques à l'économie.
2. BIOMstat for Windows: basic statistical analysis programs for use with the 3rd edition of the text Biometry by Sokal and Rohlf. Version 3.3 (Win95/98/NT/2000)
3. Chameleon Statistics: cluster analysis and data visualization
4. NTSYSpc: numerical taxonomy system for cluster and ordination analysis. Version 2.1 (Win95/98/NT/2000)
5. NTSYSpc: numerical taxonomy system for cluster and ordination analysis. Version 2.1 (Win95/98/NT/2000)
6. Genstat:a very powerful general statistics package.
7. EZ-Stat: statistical analysis program (Win)
8. SYN-TAX 2000: data analysis in ecology and systematics (Windows, Mac)
9. Fractal-D: estimate the fractal dimension of outlines from digitized images (DOS)
10. ProStat: advanced statistical analysis (Win 3.1 /Win95/98/NT)
11. SIMSTAT for Windows: powerful statistical data-analysis program (Windows)
12. NLREG: nonlinear regression and curve fitting (Windows)
13. GS+: spatial statistics for the environmental sciences, ver. 5.3 (Windows)
14. KWIKSTAT & WINKS: statistical analysis programs (DOS & Windows)
15. MVSP: multivariate statistical package (Windows 3.1/Win95/98/NT/2000)
16. Oriana: circular statistics for Windows (Windows)
17. Resampling Stats: randomization and bootstrap statistics (Win & Mac)
18. SAAP: spatial autocorrelation analysis (DOS).
19. Statistical Calculator: a programmable and extensible system for data analysis (DOS)
20. BIOMlab: learn basic statistical concepts through sampling experiments (DOS)

Une mention particulière pour Stat-200 : quoique bridé en version demo, il permet de se faire une idée très précise de l'emploi des

tests, sur un mode assez intuitif et semblable à une présentation de type « Excel ». Une grande richesse de tests et d'options. On peut montrer, par une astuce très simple, qu'il est possible, sans employer de « crack » de passer outre la limitation du logiciel, par le remploi judicieux de certains fichiers au format **txt**. A bon entendeur, salut !

TABLES

23 tables spécifiques à chaque test sont incluses dans la feuille macro ; elles peuvent être copiées en les sélectionnant dans une liste déroulante ; pour désactiver l'option, il faut choisir la ligne « **no table** » qui permet de choisir le test que l'on veut étudier :



- Mann and Whitney à 5% [d'après Jacobson, J.E., Journ. of the Amer. Stat. Ass., 1963, 1086]
- Mann and Whitney à 1%

- Fisher Yates Terry (test C1) [d'après Fisher et Yates, *Statistical tables for biological, agricultural and medical research* (Oliver and Boyd, Edinburgh)]

- limite supérieure de Fisher Yates Terry C1 à 5% et à 1% [d'après Klotz J.H., *On the normal scores two sample rank test*, J. Amer. Stat. Ass., 1964, 652-664]

-test T de Wilcoxon pour séries appariées [simplifiée, d'après Gibbons (1971) ; E. L. Lehmann (in *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco, Holden Day, 1975).]

- coefficient de corrélation r' de Spearman [simplifiée, d'après Best et Roberts, 1975]

-Friedman avec k = 3, k = 4 et k = 5
[d'après Siegel 1956, cf. http://cons-dev.univ-lyon1.fr/Enseignement/Stat/Tables/Tables.html A noter que la table donnée par ce site est incomplète ; nous avons trouvé les valeurs correspondant à k = 3 (suivant n > 8), k = 4 (suivant n > 4) et enfin k = 5 (n de 3 à 6) à la page suivante : http://www-class.unl.edu/psycrs/handcomp/hcfried.PDF.]

- Durbin Watson de m = 1 à m = 5 [régresseurs ou variables explicatives] à 5% et à 1% [J. Durbin and G.S. Watson, *Testing for serial correlation in least squares regression*, Biometrika, 1951, 30, 159-178]. A noter qu'il existe une « *coquille* » dans la table à 1% à l'intersection n = 17 et m (k) = 4 où 1.3 doit être remplacé par 1.63.

- tables de Kolmogorov Smirnov [6 tables, *Table of percentage points of Kolmogorov statistics*, Journal of the American Statistical Association Vol. 51, pp. 111 – 121).].

- table de probabilités binomiales cumulées pour le test des signes [Sign Test]

- table du test des séries de Wald Wolfowitz - run test -
[d'après : Frieda S. Swed, C. Eisenhart, *Tables for testing randomness of grouping in a sequence of alternatives*, Annals of Mathematical Statistics, 14, 1943, 66-87]

- table des valeurs critiques du test de Jonckheere - Terpstra.
[source citée supra].
- table des valeurs critiques du test de Page [in http://www.dsa.unipr.it/soliani/]
- table des valeurs critiques du coefficient de concordance de Kendall W [idem]

Ces tables donnent automatiquement les valeurs critiques dans les tableaux des résultats mais peuvent aussi être « appelées » sous forme de fichiers copiés dans une feuille d'Excel.

A suivre...

